

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|  |           |   |
|--|-----------|---|
| <p>(51) International Patent Classification<sup>6</sup> :<br/>C07H 21/04, A61K 39/12, 39/245, C12Q<br/>1/70, 1/68, G01N 33/53, C12P 1/00,<br/>C12N 5/00, A61K 39/00, 39/395</p>  | <p>A1</p> | <p>(11) International Publication Number: WO 97/27208<br/>(43) International Publication Date: 31 July 1997 (31.07.97)</p>  |
| <p>(21) International Application Number: PCT/US97/01442<br/>(22) International Filing Date: 28 January 1997 (28.01.97)<br/>(30) Priority Data:<br/>08/592,963 29 January 1996 (29.01.96) US<br/>08/757,669 29 November 1996 (29.11.96) US<br/>(71) Applicant: THE TRUSTEES OF COLUMBIA UNIVERSITY<br/>IN THE CITY OF NEW YORK [US/US]; West 116th Street<br/>and Broadway, New York, NY 10027 (US).<br/>(72) Inventors: CHANG, Yuan; Apartment 3J, 90 Morningside<br/>Drive, New York, NY 10027 (US). BOHENZKY, Roy, A.;<br/>Apartment 115, 870 East El Camino Real, Mountain View,<br/>CA 94040 (US). RUSSO, James, J.; Apartment 25E, 60<br/>Haven Avenue, New York, NY 10032 (US). EDELMAN,<br/>Isidore, S.; Apartment 61, 464 Riverside Drive, New York,<br/>NY 10027 (US). MOORE, Patrick, S.; Apartment 3J, 90<br/>Morningside Drive, New York, NY 10027 (US).<br/>(74) Agent: WHITE, John, P.; Cooper &amp; Dunham L.L.P., 1185<br/>Avenue of the Americas, New York, NY 10036 (US).</p> |           | <p>(81) Designated States: AU, CA, JP, MX, European patent (AT,<br/>BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,<br/>NL, PT, SE).<br/><br/>Published<br/>With international search report.</p> |

(54) Title: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND USES THEREOF

(57) Abstract

This invention provides an isolated nucleic acid molecule which encodes Kaposi's Sarcoma-Associated Herpesvirus (KSHV) polypeptides. This invention provides an isolated polypeptide molecule of KSHV. This invention provides an antibody specific to the polypeptide. Antisense and triplex oligonucleotide molecules are also provided. This invention provides a vaccine for Kaposi's Sarcoma (KS). This invention provides methods of vaccination, prophylaxis, diagnosis and treatment of a subject with KS and of detecting expression of a DNA virus associated with Kaposi's Sarcoma in a cell.

Applicants: Yuan Chang, et al.  
Serial No. : 09/607,179  
Filed: June 29, 2000  
Exhibit 7

BACKGROUND OF THE INVENTION

Kaposi's sarcoma-associated herpesvirus (KSHV) is a new human herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) [1,2].

Kaposi's sarcoma is the most common neoplasm occurring in persons with acquired immunodeficiency syndrome (AIDS). Approximately 15-20% of AIDS patients develop this neoplasm which rarely occurs in immunocompetent individuals. Epidemiologic evidence suggests that AIDS-associated KS (AIDS-KS) has an infectious etiology. Gay and bisexual AIDS patients are approximately twenty times more likely than hemophiliac AIDS patients to develop KS, and KS may be associated with specific sexual practices among gay men with AIDS. KS is uncommon among adult AIDS patients infected through heterosexual or parenteral HIV transmission, or among pediatric AIDS patients infected through vertical HIV transmission. Agents previously suspected of causing KS include cytomegalovirus, hepatitis B virus, human papillomavirus, Epstein-Barr virus (EBV), human herpesvirus 8, human immunodeficiency virus (HIV), and Mycoplasma penetrans. Non-infectious environmental agents, such as nitrite inhalants, also have been proposed to play a role in KS tumorigenesis. Extensive investigations, however, have not demonstrated an etiologic association between any of these agents and AIDS-KS.

UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND  
USES THEREOF

5 The invention disclosed herein was made with  
Government support under a co-operative agreement  
CCU210852 from the Centers for Disease Control and  
Prevention, and under National Institutes of Health,  
10 National Cancer Institute award CA67391 of the  
Department of Health and Human Services. Accordingly,  
the U.S. Government has certain rights in this  
invention.

15 This application is a PCT International Application  
claiming priority of U.S. Serial No. 08/757,669, filed  
November 29, 1996 and of U.S. Serial No. 08/592,963,  
filed January 29, 1996, which is a continuation-in-  
part application of PCT International Application No.  
20 PCT/US95/15138, filed November 21, 1995 and  
PCT/US95/10194, filed August 11, 1995, claiming  
priority of U.S. Serial No. 08/420,233, filed April  
11, 1995 and of U.S. Serial No. 08/343,101, filed  
November 21, 1994, which is a continuation-in-part of  
25 U.S. Serial No. 08/292,363, filed August 18, 1994,  
which is hereby incorporated by reference.

Throughout this application, various publications may  
be referenced by Arabic numerals in brackets. Full  
30 citations for these publications may be found at the  
end of the Detailed Description of the Invention. The  
disclosures of all publications cited herein are in  
their entirety hereby incorporated by reference into  
this application to more fully describe the state of  
35 the art to which this invention pertains.



## BRIEF DESCRIPTION OF THE FIGURES

### Figure 1:

5 Annotated long unique region (LUR) and terminal repeat (TR) of the KSHV genome. The orientation of identified ORFs in the LUR are denoted by the direction of arrows, with ORFs similar to HSV in dark blue and dis-similar ORFs in light blue. Seven blocks (numbered) of conserved herpesvirus genes with nonconserved interblock regions (lettered) are shown under the kilobase marker; the block numbering scheme differs from the original description by Chee (Chee et al., 1990, *Curr. Topics Microbiol. Immunol.* 154, 125-169).  
10 The overlapping cosmid (Z prefix) and lambda (L prefix) clones used to map the KSHV genome are compared to the KS5 lambda phage clone from a KS lesion and shown below. Features and putative coding regions not specifically designated are shown above the ORF map. Repeat regions are shown as white lines (frnk, vnct, waka/jwka, zppa, moi, mdsk). Putative coding regions and other features (see Experimental Details Section I) not designated as ORFs are shown as solid lines.  
20  
25

### Figure 2A-2D:

(Fig. 2A) Sequence of terminal repeat unit (TR) demonstrating its high G-C content (SEQ ID NO:16). Sequences highly similar to conserved herpesvirus pac1 sites are underlined with less similar sites to specific pac1 and pac2 sequences italicized. (Fig. 2B) Southern blot of DNA from BC-1 (lane 1), BCP-1 (lane 2) and a KS lesion (lane 3) digested with NdeII which cuts once in the TR sequence and probed with a plasmid containing the TR sequence. The intense  
30  
35

SUMMARY OF THE INVENTION

This invention provides an isolated nucleic acid molecule which encodes Kaposi's Sarcoma-Associated Herpesvirus (KSHV) polypeptides. This invention  
5 provides an isolated polypeptide molecule of KSHV. This invention provides an antibody specific to the polypeptide. Antisense and triplex oligonucleotide molecules are also provided. This invention provides  
10 a vaccine for Kaposi's Sarcoma (KS). This invention provides methods of vaccination, prophylaxis, diagnosis and treatment of a subject with KS and of detecting expression of a DNA virus associated with Kaposi's sarcoma in a cell.

15

Both KSHV MIP genes encode 19 residue N-terminus hydrophobic secretory leader sequences which are relatively poorly conserved (vMIP-I also has a second C-C dimer in the hydrophobic leader sequence without similarity to the chemokine dicysteine motif). Potential C-linked glycosylation sites for vMIP-I (gapped positions 22 and 27) are not present in vMIP-II, which has only one predicted potential serine glycosylation site (position 51) not found in vMIP-I. Fig. 3B. Alignment of the KSHV vIL-6 to human IL-6. Fig. 3C-1 and 3C-2. Alignment of the KSHV vIRF polypeptide to human ICSBP and ISGF3 with the putative ICS-binding typtophans (W) for ICSBP and ISGF3 in italics.

#### Figures 4A-4F:

Northern hybridization of total RNA extracted from BCP-1 and BC-1 cells with or without 48 hour incubation with TPA and control PBHR1 cells after TPA incubation. All four genes (Fig. 4A, vMIP-I; Fig. 4B, vMIP-II; Fig. 4C, vIL-6; Fig. 4D, vIRF) are TPA inducible but constitutive, noninduced expression of vIL-6 (Fig. 4C) and vIRF (Fig. 4D) is also evident for BCP-1 and BC-1 and of vMIP-I for BCP-1 (Fig. 4A). Representative hybridizations to a human  $\beta$ -actin probe (Figs. 4E-4F) demonstrate comparable loading of RNA for cell preparations.

#### Figures 5A-5B:

Fig. 5A. Immunoblot of rabbit anti-peptide antibodies generated from amino acid sequences of vIL-6, THYSPPKFDR (SEQ ID NO:2) and PDVTPDVHDR (SEQ ID NO:3), against cell lysates of BCP-1, BC-1, PBHR1 cell lines with and without TPA induction (lanes 1-6), 1  $\mu$ g human rIL-6 (lane 7),

hybridization band at 0.8 kb represents multiple copies of the NdeII-digested single unit TR (Fig. 2C). A schematic representation (Fig. 2C) of genome structures of KSHV in BCP-1 and BC-1 cell lines consistent with the data presented in (Fig. 2B) and (Fig. 2D). TagI (T) sites flank the TR regions and Nde II (N) sites are within the TRs. Lower case tr refers to the deleted truncated TR unit at the left end of the unique region. DR represents the duplicated region of the LUR buried within the TR. (Fig. 2D) Southern blot hybridization with TR probe of DNA from BC-1 (lane 1), BCP-1 (lane 2), a KS lesion (lane 3), and HBL-6 (lane 4) digested with Tag I, which does not cut in the TR. Tag I-digested DNA from both BC-1 (lane 1) and HBL-6 (lane 4) show similar TR hybridization patterns suggesting identical insertion of a unique sequence into the TR region, which sequencing studies demonstrate is a duplicated portion of the LUR (see Experimental Details Section). BCP-1 TR hybridization (lane 2) shows laddering consistent with a virus population having variable TR region lengths within this cell line due to lytic replication. The absence of TR laddering in KS lesion DNA (lane 3) suggests that a clonal virus population is present in the tumor.

#### Figures 3A-3C:

CLUSTAL W alignments of KSHV-encoded polypeptide sequences to corresponding human cell signaling pathway polypeptide sequences. Fig. 3A. Two KSHV MIP-like polypeptides (vMIP-I and vMIP-II) are compared to human MIP-1 $\alpha$ , MIP-1 $\beta$  and PANTES (amino acid identity to vMIP-I indicated by black reverse shading, to vMIP-II alone by gray reverse shading, and the C-C dimer motif is italicized).

demonstrates vIL-6 production in both KSHV-infected cell lines and tissues. The KSHV-infected cell line BCP-1 (Fig. 7A), but not the control EBV-infected cell line F3HRL (Fig. 7B), shows prominent cytoplasmic vIL-6 localization. (Fig. 7C) Cytoplasmic localization of vIL-6 in spindle-shaped cells from an AIDS-KS lesion. Of eight KS lesions, only one had readily identifiable vIL-6 staining of a subpopulation of cells. In contrast, the majority of pelleted lymphoma cells from a nonAIDS, EBV-negative PEL have intense vIL-6 staining (Fig. 7E). No immunostaining is present in control angiosarcoma (Fig. 7D) or multiple myeloma tissues (Fig. 7F).

#### Figures 8A-8D:

Double antibody labeling of anti-vIL-6 and cell surface antigens. Examples of both CD34 and CD20 colocalization with vIL-6 were found in a KS lesion. Fig. 8A. CD34 (red) and vIL-6 colocalize (blue) in a KS spindle cell (arrow). Purple coloration is due to overlapping chromagen staining (100X). Fig. 8B. CD45 common leukocyte antigen staining (blue, arrow) on vIL-6 (red) expressing Kaposi's sarcoma cells (100X). Fig. 8C. Low power magnification (20X) demonstrating numerous vIL-6 producing hematopoietic cells (red) in a lymph node from a patient with KS. Arrows only indicate the most prominently staining cells; nuclei counterstained with hematoxylin. Fig. 8D. Colocalization of CD20 (brown, arrows) with vIL-6 (red) in an AIDS-KS patient's lymph node (100X).

35

#### Figure 9:

and concentrated COS7 rvIL-6 and r6-LIV supernatants (lanes 8-9). Anti-vIL-6 antibodies specifically recognize the viral IL-6 polypeptide in both recombinant supernatants and cell lines but not human IL-6. The BCP-1 cell line constitutively expresses low levels of vIL-6 whereas polypeptide expression increases on TPA treatment for both BC-1 (KSHV and EBV coinfecting) and BCP-1 (KSHV infection alone) indicating lytic phase expression. Preimmune sera from immunized rabbits did not react on immunoblotting to any of the preparations. Fig. 5B. Anti-huIL-6 monoclonal antibodies do not cross-react with cell-associated or recombinant vIL-6 preparations.

#### Figure 6:

Dose-response curves for <sup>3</sup>H-thymidine uptake in IL-6-dependent B9 mouse plasmacytoma cells with serial dilutions of rhvIL-6 (filled squares) and COS7 supernatants of rvIL-6 (filled circles), r6-LIV (open squares) or control LacZ (open circles) pMET7 transfections. Undiluted rvIL-6 supernatants from this transfection lot show similar B9 proliferation activity to huIL-6 >0.02 ng/ml whereas the reverse construct (r6-LIV) and the LacZ control show no increased ability to induce B9 proliferation. Concentrated supernatants at greater than 1:1 dilution may have increased activity due to concentration of COS7 conditioning factors.

#### Figures 7A-7F:

Rabbit anti-vIL-6 peptide antibody reactivity localized using goat-antirabbit immunoglobulin-peroxidase conjugate (brown) with hematoxylin counterstaining (blue) at X100 magnification

DETAILED DESCRIPTION OF THE INVENTIONDefinitions

5 The following standard abbreviations are used throughout the specification to indicate specific nucleotides:

|    |             |             |
|----|-------------|-------------|
|    | C=cytosine  | A=adenosine |
| 10 | T=thymidine | G=guanosine |

The term "nucleic acid", as used herein, refers to either DNA or RNA, including complementary DNA (cDNA), genomic DNA and messenger RNA (mRNA). As used herein,  
15 "genomic" means both coding and non-coding regions of the isolated nucleic acid molecule. "Nucleic acid sequence" refers to a single- or double- stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It includes both  
20 self-replicating plasmids, infectious polymers of DNA or RNA and nonfunctional DNA or RNA.

The term "polypeptide", as used herein, refers to either the full length gene product encoded by the  
25 nucleic acid, or portions thereof. Thus, "polypeptide" includes not only the full-length protein, but also partial-length fragments, including peptides less than fifty amino acid residues in length.

30 The term "SSC" refers to a citrate-saline solution of 0.15 M sodium chloride and 20 mM sodium citrate. Solutions are often expressed as multiples or fractions of this concentration. For example, 6XSSC  
35 refers to a solution having a sodium chloride and sodium citrate concentration of 6 times this amount or 0.9 M sodium chloride and 120 mM sodium citrate.

Quantification of CCC/CD4 cell infection by  
primary NSI SF162 and M23 HIV-1 strains and HIV-2  
strain ROD/B in the presence or absence of  
vMIP-I. CCC/CD4 cells were transiently  
5 cotransfected with CCR5 alone, CCR5 plus empty  
pMET7 vector, CCR5 plus vMIP-I in pMET7 vector,  
or CCR5 plus the reverse orientation I-PIMv. The  
results after 72 hours of incubation with each  
retrovirus are expressed as a percentage of the  
10 foci forming units for cells transfected with  
CCR5 alone. The forward vMIP-I construct  
inhibited NSI HIV-1 replication but not HIV-2  
replication while the reverse I-PIMv construct  
had no effect on replication of any of the  
15 retroviruses.



length sequences. It being further understood that the sequence includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in a specific host cell.

5 A nucleic acid probe is "specific" for a target organism of interest if it includes a nucleotide sequence which when detected is determinative of the presence of the organism in the presence of a  
10 heterogeneous population of proteins and other biologics. A specific nucleic acid probe is targeted to that portion of the sequence which is determinative of the organism and will not hybridize to other sequences, especially those of the host, where a  
15 pathogen is being detected.

The phrase "expression cassette", refers to nucleotide sequences which are capable of affecting expression of a structural gene in hosts compatible with such  
20 sequences. Such cassettes include at least promoters and optionally, transcription termination signals. Additional factors necessary or helpful in effecting expression may also be used as described herein.

25 The term "operably linked" as used herein refers to linkage of a promoter upstream from a DNA sequence such that the promoter mediates transcription of the DNA sequence.

30 The term "vector", refers to viral expression systems, autonomous self-replicating circular DNA (plasmids), and includes both expression and nonexpression plasmids. Where a recombinant microorganism or cell culture is described as hosting an "expression  
35 vector," this includes both extrachromosomal circular DNA and DNA that has been incorporated into the host chromosome(s). Where a vector is being maintained by

0.2XSSC refers to a solution 0.2 times the SSC concentration or 0.03 M sodium chloride and 4 mM sodium citrate.

5 The phrase "selectively hybridizing to" and the phrase "specific hybridization" describe a nucleic acid probe that hybridizes, duplexes or binds only to a particular target DNA or RNA sequence when the target sequences are present in a preparation of total  
10 cellular DNA or RNA. By selectively hybridizing it is meant that a probe binds to a given target in a manner that is detectable in a different manner from non-target sequence under high stringency conditions of hybridization.

15 "Complementary" or "target" nucleic acid sequences refer to those nucleic acid sequences which selectively hybridize to a nucleic acid probe. Proper annealing conditions depend, for example, upon a  
20 probe's length, base composition, and the number of mismatches and their position on the probe, and must often be determined empirically. For discussions of nucleic acid probe design and annealing conditions, see, for example, Sambrook et al. (1989) Molecular  
25 Cloning: A Laboratory Manual (2nd ed.), Cold Spring Harbor Laboratory, Vols. 1-3 or Ausubel, F., et al. (1987) Current Protocols in Molecular Biology, New York.

30 The phrase "nucleic acid molecule encoding" refers to a nucleic acid molecule which directs the expression of a specific polypeptide. The nucleic acid sequences include both the DNA strand sequence that is transcribed into RNA, the complementary DNA strand,  
35 and the RNA sequence that is translated into protein. The nucleic acid molecule includes both the full length nucleic acid sequence as well as non-full

Waterman (1981) Adv. Appl. Math. 2:482, by the algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443, by the search-for-similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. 85:2444, or  
5 by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in GCG, the Wisconsin Genetics Software Package Release 8.0, Genetics Computer Group, 575 Science Dr., Madison, WI).

10 As applied to polypeptides, the terms "substantial identity" or "substantial sequence identity" mean that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap which share at least 90 percent sequence identity,  
15 preferably at least 95 percent sequence identity, more preferably at least 99 percent sequence identity or more.

"Percentage amino acid identity" or "percentage amino acid sequence identity" refers to a comparison of the amino acids of two polypeptides which, when optimally aligned, have approximately the designated percentage of the same amino acids. For example, "95% amino acid identity" refers to a comparison of the amino acids of  
20 two polypeptides which when optimally aligned have 95% amino acid identity. Preferably, residue positions which are not identical differ by conservative amino acid substitutions. For example, the substitution of amino acids having similar chemical properties, such  
25 as charge or polarity, are not likely to effect the properties of a protein. Examples include glutamine for asparagine or glutamic acid for aspartic acid.  
30

The phrase "substantially purified" or "isolated" when  
35 referring to a herpesvirus polypeptide, means a chemical composition which is essentially free of other cellular components. It is preferably in a

a host cell, the vector may either be stably replicated by the cells during mitosis as an autonomous structure, or is incorporated within the host's genome.

5

The term "plasmid" refers to an autonomous circular DNA molecule capable of replication in a cell, and includes both the expression and nonexpression types. Where a recombinant microorganism or cell culture is described as hosting an "expression plasmid", this includes latent viral DNA integrated into the host chromosome(s). Where a plasmid is being maintained by a host cell, the plasmid is either being stably replicated by the cells during mitosis as an autonomous structure or is incorporated within the host's genome.

10

15

The phrase "recombinant protein" or "recombinantly produced protein" refers to a polypeptide produced using non-native cells. The cells produce the protein because they have been genetically altered by the introduction of the appropriate nucleic acid sequence.

20

25

The following terms are used to describe the sequence relationships between two or more nucleic acid molecules: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing or may comprise a complete cDNA or gene sequence.

30

35

Optimal alignment of sequences in a comparison window may be conducted by the algorithm of Smith and

"Biological sample" as used herein refers to any sample obtained from a living organism or from an organism that has died. Examples of biological samples include body fluids and tissue specimens.

5

It will be readily understood by those skilled in the art and it is intended here, that when reference is made to particular sequence listings, such reference includes sequences which substantially correspond to the listing and it's complement, including allowances for minor sequencing errors, single base changes, deletions, substitutions and the like, such that any such sequence variation corresponds to the nucleic acid sequence of the pathogenic organism or disease marker to which the relevant sequence listing relates.

10

15

#### 1. Nucleic Acid Molecule from KSHV

This invention provides an isolated nucleic acid molecule which encodes a Kaposi's sarcoma-associated herpesvirus (KSHV) polypeptide.

20

In one embodiment, the isolated nucleic acid molecule which encodes a KSHV polypeptide has the nucleotide sequence as set forth in GenBank Accession Number U75698 and the start and stop codons set forth in Table 1. In another embodiment, the isolated nucleic acid molecule which encodes a KSHV polypeptide has the amino acid sequence defined by the translation of the nucleotide sequence set forth in GenBank Accession Number U75698 and the start and stop codons set forth in Table 1.

25

30

In one embodiment, the isolated nucleic acid molecule for a KSHV polypeptide has the 5' untranslated sequence as set forth in GenBank Accession Number U75698 upstream of the ATG start codon. In another

35

homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. Generally, a substantially purified or isolated protein will comprise more than 80% of all macromolecular species present in the preparation. Preferably, the protein is purified to represent greater than 90% of all macromolecular species present. More preferably the protein is purified to greater than 95%, and most preferably the protein is purified to essential homogeneity, wherein other macromolecular species are not detected by conventional techniques.

The phrase "specifically binds to an antibody" or "specifically immunoreactive with", when referring to a polypeptide, refers to a binding reaction which is determinative of the presence of the KSHV polypeptide of the invention in the presence of a heterogeneous population of polypeptides and other biologics including viruses other than KSHV. Thus, under designated immunoassay conditions, the specified antibodies bind to the KSHV antigen and do not bind in a significant amount to other antigens present in the sample.

"Specific binding" to an antibody under such conditions may require an antibody that is selected for its specificity for a particular antigen. For example, antibodies raised to KSHV antigens described herein can be selected to obtain antibodies specifically immunoreactive with KSHV polypeptides and not with other polypeptides.

for 2 hours at 37° C prior to infection. Infected cells are observed by demonstrating morphological changes, as well as being viral antigen positive.

5 For KSHV isolation, the virus is either harvested directly from cell culture fluid by centrifugation, or the infected cells are harvested, homogenized or lysed and the virus is separated from cellular debris and purified by standard methods of isopycnic sucrose  
10 density gradient centrifugation.

One skilled in the art may isolate and propagate KSHV employing the following protocol. Long-term establishment of a B lymphoid cell line infected with  
15 KSHV (e.g., RCC-1, HBL-6 or BCBL-1) is accomplished using body-cavity based lymphomas and standard techniques (Glick, 1980, *Fundamentals of Human Lymphoid Culture*, Marcel Dekker, New York; Knowles et al., 1989, *Blood* 73, 792-798; Metcalf, 1984, *Clonal Culture of Hematopoietic Cells: Techniques and Applications*, Elsevier, New York).  
20

Fresh lymphoma tissue containing viable infected cells is filtered to form a single cell suspension. The  
25 cells are separated by Ficoll-Plaque centrifugation and lymphocyte layer is removed. The lymphocytes are then placed at  $>1 \times 10^5$  cells/ml into standard lymphocyte tissue culture medium, such as RPMI 1640 supplemented with 10% fetal calf serum. Immortalized lymphocytes  
30 containing KSHV are indefinitely grown in the culture media while non-immortalized cells die during course of prolonged cultivation.

Further, KSHV may be propagated in a new cell line by removing media supernatant containing the virus from  
35 a continuously-infected cell line at a concentration of  $>1 \times 10^5$  cells/ml. The media is centrifuged at 2000xg

embodiment, the isolated nucleic acid molecule for a KSHV polypeptide has the 3' untranslated sequence as set forth in GenBank Accession Number U75696 downstream of the stop codon.

5 In one embodiment the isolated nucleic acid molecule is genomic DNA. In another embodiment the isolated nucleic acid molecule is cDNA. In another embodiment RNA is derived from the isolated nucleic acid molecule or is capable of hybridizing with the isolated nucleic acid molecule.

10 Further, the nucleic acid molecule above may be associated with lymphoproliferative diseases including, but not limited to: Hodgkin's disease, non-Hodgkin's lymphoma, lymphatic leukemia, lymphosarcoma, splenomegaly, reticular cell sarcoma, Sezary's syndrome, mycosis fungoides, central nervous system lymphoma, AIDS related central nervous system lymphoma, post-transplant lymphoproliferative disorders, and Burkitt's lymphoma. A lymphoproliferative disorder is characterized as being the uncontrolled clonal or polyclonal expansion of lymphocytes involving lymph nodes, lymphoid tissue and other organs.

#### A. Isolation and Propagation of KSHV

30 KSHV can be propagated in vitro. For example, techniques for growing herpesviruses have been described by Ablashi et al. in Virology 184, 545-552. Briefly, PHA stimulated cord blood mononuclear cells, macrophage, neuronal, or glial cell lines are cocultivated with cerebrospinal fluid, plasma, peripheral blood leukocytes, or tissue extracts containing viral infected cells or purified virus. The recipient cells are treated with 5 ug/ml polybrene



The method for isolating the KSHV genome is based on Pellicer et al., 1978, Cell 14, 133-141 and Gibson and Roizmann, 1972, J. Virol. 10, 1044-52.

5 A final method for isolating the KSHV genome is clamped homogeneous electric field (CHEF) gel electrophoresis. Agarose plugs are prepared by resuspending cells infected with KSHV in 1% LMP agarose (Biorad) and 0.9% NaCl at 42°C to a final  
10 concentration of  $2.5 \times 10^7$  cells/ml. Solidified agarose plugs are transferred into lysis buffer (0.5M EDTA pH 8.0, 1% sarcosyl, proteinase K at 1 mg/ml final concentration) and incubated for 24 hours. Approximately  $10^7$  cells are loaded in each lane. Gels  
15 are run at a gradient of 6.0 V/cm with a run time of 28 h on a CHEF Mapper XA pulsed field gel electrophoresis apparatus (Biorad), Southern blotted and hybridized to KS631Bam, KS330Bam and an EBV terminal repeat sequence.

20 To make a new cell line infected with KSHV, already-infected cells are co-cultivated with a Raji cell line separated by a 0.45 $\mu$  filter. Approximately,  $1-2 \times 10^6$  already-infected BCBL-1 and  $2 \times 10^6$  Raji cells are co-  
25 cultivated for 2-20 days in supplemented RPMI alone or with 20 ng/ml 12-O-tetradecanoyl phorbol-13-acetate (TPA). After 2-20 days co-cultivation, Raji cells are removed, washed and placed in supplemented RPMI 1640 media. A Raji culture co-cultivated with BCBL-1 in 20  
30 ng/ml TPA for 2 days survived and has been kept in continuous suspension culture for >10 weeks. This cell line, designated RCC-1 (Raji Co-Culture, No.1) remains PCR positive for the KSHV sequence after multiple passages. RCC-1 cells periodically undergo  
35 rapid cytolysis suggestive of lytic reproduction of KSHV. Thus, RCC-1 is a Raji cell line newly-infected with KSHV.

for 10 minutes and filtered through a 0.45 $\mu$  filter to remove cells. The media is applied in a 1:1 volume with cells growing at  $>1 \times 10^6$  cells/ml for 48 hours. The cells are washed, pelleted and placed in fresh culture medium, then tested for KSHV after 14 days.

KSHV may be isolated from a cell line in the following manner. An infected cell line is lysed using standard methods, such as hyposmotic shock or Dounce homogenization or using repeated cycles of freezing and thawing in a small volume ( $<3$  ml), and pelleted at 2000xg for 10 minutes. The supernatant is removed and centrifuged again at 10,000xg for 15 minutes to remove nuclei and organelles. The resulting low-speed, cell-free supernatant is filtered through a 0.45 $\mu$  filter and centrifuged at 100,000xg for 1 hour to pellet the virus. The virus can then be washed and re-pelleted. The DNA is extracted from the viral pellet by standard techniques (e.g., phenol/ chloroform) and tested for the presence of KSHV by Southern blotting and/or PCR using the specific probes described above.

For banding whole virion, the low-speed cell-free supernatant is adjusted to contain 7% PEG-8000. The PEG-supernatant is spun at 10,000 xg for 30 min. The supernatant is poured off and the pellet collected and resuspended in a small volume (1-2 ml) of virus buffer (VB, 0.1 M NaCl, 0.01 M Tris, pH 7.5). The virion are isolated by centrifugation at 25,000 rpm in a 10-50% sucrose gradient made with VB. One ml fractions of the gradient are obtained by standard techniques (e.g., using a fractionator) and each fraction is tested by dot blotting using specific hybridizing probes to determine the gradient fraction containing the purified virus (preparation of the fraction is needed in order to detect the presence of the virus, i.e., standard DNA extraction).

In one embodiment the molecule is DNA. In another embodiment the molecule is RNA.

5 In one embodiment the TR molecule contains cis-active elements required for DNA replication and packaging. In another embodiment the TR molecule is contained in a gene-cloning vector. In another embodiment the TR molecule is contained in a gene-therapy vector. In another embodiment the gene-therapy vector is  
10 expressed in lymphoid cells. In another embodiment, the TR comprises a molecular marker for determining the clonality of a tumor. In another embodiment, the marker provides a defining feature of the natural history of a tumor in a diagnostic assay.

15 This invention provides a B-lymphotrophic DNA vector comprising a plasmid or other self-replicable DNA molecule containing the 801 bp KSHV TR or a portion thereof.

20

High stringency hybridization conditions are selected at about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the salt  
25 concentration is at least about 0.02 molar at pH 7 and the temperature is at least about 60°C. As other factors may significantly affect the stringency of hybridization, including, among others, base composition and size of the complementary strands, the presence of organic solvents, i.e. salt or formamide concentration, and the extent of base mismatching, the combination of parameters is more important than the  
30 absolute measure of any one. For example, high stringency may be attained by overnight hybridization at about 68°C in a 6X SSC solution, washing at room

35

RCC-1 and RCC-1<sub>PH</sub> were deposited on October 19, 1994 under ATCC Accession No. CRL 11734 and CRL 11735, respectively, pursuant to the Budapest Treaty on the International Deposit of Microorganisms for the Purposes of Patent Procedure with the Patent Culture Depository of the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852 U.S.A. HBL-6 was deposited (as BHL-6) on November 18, 1994 under ATCC Accession No. CRL 11762 pursuant to the Budapest Treaty on the International Deposit of Microorganisms for the Purposes of Patent Procedure with the Patent Culture Depository of the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852 U.S.A.

15

#### B. Hybridization Probes of KSHV

This invention provides a nucleic acid molecule of at least 14 nucleotides capable of specifically hybridizing with the isolated nucleic acid molecule as set forth in GenBank Accession Numbers U75698, U75699, U75700.

In one embodiment the nucleic acid molecule set forth in GenBank Accession Number U75698 comprises the long unique region (LUR) encoding KSHV polypeptides. In another embodiment the nucleic acid molecule set forth in GenBank Accession Number U75699 comprises the prototypical terminal repeat (TR). In another embodiment the nucleic acid molecule set forth in GenBank Accession Number U75700 comprises the incomplete terminal repeat (ITR).

In one embodiment the molecule is 8 to 36 nucleotides. In another embodiment the molecule is 12 to 25 nucleotides. In another embodiment the molecule is 14 nucleotides.

it contains an upstream promoter in the presence of the appropriate RNA polymerase.

As defined herein nucleic acid probes may be DNA or RNA fragments. DNA fragments can be prepared, for example, by digesting plasmid DNA, or by use of PCR, or synthesized by either the phosphoramidite method described by Beaucage and Carruthers, 1981, Tetrahedron Lett. 22, 1859-1862 or by the triester method according to Matteucci et al., 1981, Am. Chem. Soc. 103:3185. A double stranded fragment may then be obtained, if desired, by annealing the chemically synthesized single strands together under appropriate conditions or by synthesizing the complementary strand using DNA polymerase with an appropriate primer sequence. Where a specific sequence for a nucleic acid probe is given, it is understood that the complementary strand is also identified and included. The complementary strand will work equally well in situations where the target is a double-stranded nucleic acid. It is also understood that when a specific sequence is identified for use as a nucleic probe, a subsequence of the listed sequence which is 25 base pairs (bp) or more in length is also encompassed for use as a probe.

The nucleic acid molecules of the subject invention also include molecules coding for polypeptide analogs, fragments or derivatives of antigenic polypeptides which differ from naturally-occurring forms in terms of the identity or location of one or more amino acid residues (deletion analogs containing less than all of the residues specified for the polypeptide, substitution analogs wherein one or more residues specified are replaced by other residues and addition analogs where in one or more amino acid residues is added to a terminal or medial portion of the

temperature with 6X SSC solution, followed by washing at about 68°C in a 0.6X SSC solution.

Hybridization with moderate stringency may be attained for example by: 1) filter pre-hybridizing and hybridizing with a solution of 3X SSC, 50% formamide, 0.1M Tris buffer at pH 7.5, 5X Denhardt's solution; 2.) pre-hybridization at 37°C for 4 hours; 3) hybridization at 37°C with amount of labeled probe equal to 3,000,000 cpm total for 16 hours; 4) wash in x SSC and 0.1% SDS solution; 5) wash 4X for 1 minute each at room temperature in 4X SSC at 60°C for 30 minutes each; and 6) dry and expose to film.

Nucleic acid probe technology is well known to those skilled in the art who readily appreciate that such probes may vary greatly in length and may be labeled with a detectable label, such as a radioisotope or fluorescent dye, to facilitate detection of the probe. DNA probe molecules may be produced by insertion of a DNA molecule having the full-length or a fragment of the isolated nucleic acid molecule of the DNA virus into suitable vectors, such as plasmids or bacteriophages, followed by transforming into suitable bacterial host cells, replication in the transformed bacterial host cells and harvesting of the DNA probes, using methods well known in the art. Alternatively, probes may be generated chemically from DNA synthesizers.

RNA probes may be generated by inserting the full length or a fragment of the isolated nucleic acid molecule of the DNA virus downstream of a bacteriophage promoter such as T3, T7 or SP6. Large amounts of RNA probe may be produced by incubating the labeled nucleotides with a linearized isolated nucleic acid molecule of the DNA virus or its fragment where

In another embodiment, DHFR has the amino acid sequence as set forth in SEQ ID NO:1.

5 In another embodiment, KSHV DHFR is inhibited by a sulfa drug known to inhibit bacterial DHFR. In a preferred embodiment, KSHV DHFR is inhibited by methotrexate or a derivative thereof known to inhibit mammalian DHFR. In another embodiment, the sulfa  
10 drug, methotrexate or a derivative thereof is selective among the human herpesviruses for inhibition of KSHV.

This invention provides the isolated KSHV polypeptide comprising thymidylate synthase (TS) encoded by ORF  
15 70. In one embodiment, TS participates in KSHV nucleotide metabolism. In another embodiment, TS comprises an enzyme essential for viral replication, inhibition of which prevents virus production. In another embodiment, TS comprises a subunit vaccine.  
20 In another embodiment, TS comprises an antigen for immunologic assays.

This invention provides the isolated KSHV polypeptide comprising DNA polymerase encoded by ORF 9. In one  
25 embodiment, DNA polymerase comprises an enzyme essential for viral replication, inhibition of which prevents virus production. In another embodiment, DNA polymerase comprises a subunit vaccine. In another embodiment, DNA polymerase comprises an antigen for  
30 immunologic assays.

This invention provides the isolated KSHV polypeptide comprising alkaline exonuclease encoded by ORF 37. In  
35 one embodiment, alkaline exonuclease packages KSHV DNA into the virus particle. In another embodiment, alkaline exonuclease comprises an enzyme essential for viral replication, inhibition of which prevents virus

polypeptides) and which share some or all properties of naturally-occurring forms. These molecules include: the incorporation of codons "preferred" for expression by selected non-mammalian hosts; the provision of sites for cleavage by restriction endonuclease enzymes; and the provision of additional initial, terminal or intermediate DNA sequences that facilitate construction of readily expressed vectors.

10 C. Polypeptides of KSHV and Antibodies  
(Ab's) Thereeto

This invention provides an isolated KSHV polypeptide, one from the list as set forth in Table 1 and below.

15 This invention provides the isolated KSHV polypeptide comprising viral macrophage inflammatory protein III (vMIP-III). In one embodiment, vMIP-III comprises an orphan cytokine. In another embodiment, vMIP-III is encoded by nucleotides 22,529-22,185. In another embodiment, vMIP-III comprises an anti-inflammatory drug. In a preferred embodiment, the drug is useful in treatment of an autoimmune disorder. In the most preferred embodiment, the drug is useful in treatment of rheumatoid arthritis.

25 This invention provides the isolated KSHV polypeptide comprising dihydrofolate reductase (DHFR) encoded by ORF 2. In one embodiment, DHFR participates in KSHV nucleotide synthesis. In another embodiment, DHFR comprises an enzyme essential for viral replication, inhibition of which prevents virus production. In another embodiment, DHFR comprises a subunit vaccine. In another embodiment, DHFR comprises an antigen for immunologic assays.

30

35



This invention provides the isolated KSHV polypeptide comprising viral protein kinase encoded by ORF 36. In another embodiment, viral protein kinase comprises an antigen for immunologic assays. In another  
5 embodiment, viral protein kinase comprises a subunit vaccine.

This invention provides the isolated KSHV polypeptide comprising lytic cycle transactivator protein (LCTP) encoded by ORF 50. In one embodiment, LCTP is  
10 required for activation of productive infection from the latent state. In another embodiment, LCTP is inhibited by known antiviral drugs. In another embodiment, prevention of LCTP expression maintains  
15 the virus in a latent state unable to replicate.

This invention provides the isolated KSHV polypeptide comprising ribonucleotide reductase, a two-subunit enzyme in which the small and large subunits are  
20 encoded by ORF 60 and ORF 61, respectively. In another embodiment, ribonucleotide reductase catalyzes conversion of ribonucleotides into deoxyribonucleotides for DNA replication. In another embodiment, ribonucleotide reductase is inhibited by  
25 known antiviral drugs in terminally differentiated cells not expressing cellular ribonucleotide reductase. In another embodiment, ribonucleotide reductase comprises an antigen for immunologic assays. In another embodiment, ribonucleotide reductase  
30 comprises a subunit vaccine. In another embodiment, ribonucleotide reductase comprises a transforming agent for establishment of immortalized cell lines.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF K1.  
35

production. In another embodiment, alkaline exonuclease comprises a subunit vaccine. In another embodiment, alkaline exonuclease comprises an antigen for immunologic assays.

5

This invention provides the isolated KSHV polypeptide comprising helicase-primase, subunits 1, 2 and 3 encoded by ORFs 40, 41 and 44, respectively. In one embodiment, helicase-primase comprises an enzyme activity essential for viral DNA replication. In another embodiment, helicase-primase is inhibited by nucleotide analogs. In another embodiment, helicase-primase is inhibited by known antiviral drugs. In another embodiment, inhibition of helicase-primase prevents KSHV replication.

10

15

This invention provides the isolated KSHV polypeptide comprising uracil DNA glycosylase (UDG) encoded by ORF 46. In one embodiment, uracil DNA glycosylase comprises an enzyme essential for KSHV DNA repair during DNA replication. In another embodiment, uracil DNA glycosylase is inhibited by known antiviral drugs. In another embodiment, uracil DNA glycosylase comprises a subunit vaccine. In another embodiment, uracil DNA glycosylase comprises an antigen for immunologic assays.

20

25

This invention provides the isolated KSHV polypeptide comprising single-stranded DNA binding protein (SSBP) encoded by ORF 06. In one embodiment, SSBP comprises an enzyme essential for KSHV DNA replication. In another embodiment, SSBP is inhibited by known antiviral drugs. In another embodiment, SSBP increases the processivity of polymerase reactions such as in the conventional PCR method for DNA amplification.

30

35

This invention provides the isolated KSHV polypeptide comprising vMIP-I encoded by ORF K6. In one embodiment, vMIP-I comprises an anti-inflammatory drug. In a preferred embodiment, the drug is useful  
5 in treatment of an autoimmune disorder. In the most preferred embodiment, the drug is useful in treatment of rheumatoid arthritis.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF K7.  
10

This invention provides the isolated KSHV polypeptide comprising Bcl-2 encoded by ORF 16.

This invention provides the isolated KSHV polypeptide comprising capsid protein I encoded by ORF 17.  
15

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 18.  
20

This invention provides the isolated KSHV polypeptide comprising tegument protein I encoded by ORF 19.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 20.  
25

This invention provides the isolated KSHV polypeptide comprising thymidine kinase encoded by ORF 21.

This invention provides the isolated KSHV polypeptide comprising glycoprotein H encoded by ORF 22.  
30

In one embodiment, the isolated KSHV polypeptide comprises the protein encoded by ORF 23.  
35

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 24.

This invention provides the isolated KSHV polypeptide comprising complement-binding protein (v-CBP; CCP encoded by ORF 4.

5 This invention provides the isolated KSHV polypeptide comprising transport protein encoded by ORF 7.

This invention provides the isolated KSHV polypeptide comprising glycoprotein B encoded by ORF 8.

10 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 10.

15 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 11.

This invention provides the isolated KSHV polypeptide comprising viral interleukin 6 (vIL-6) encoded by ORF K2. In one embodiment, antibodies selectively recognizing vIL-6 allow differentiation among lymphomas.

20

This invention provides the isolated KSHV polypeptide comprising BHV4-IE1 I encoded by ORF K3.

25 This invention provides the isolated KSHV polypeptide comprising vMIP-II encoded by ORF K4. In one embodiment, vMIP-II comprises an anti-inflammatory drug. In a preferred embodiment, the drug is useful in treatment of an autoimmune disorder. In the most preferred embodiment, the drug is useful in treatment of rheumatoid arthritis.

30

This invention provides the isolated KSHV polypeptide comprising BHV4-IE1 II encoded by ORF K5.

35

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 38.

5 This invention provides the isolated KSHV polypeptide comprising glycoprotein M encoded by ORF 39.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 42.

10 This invention provides the isolated KSHV polypeptide comprising capsid protein III encoded by ORF 43.

This invention provides the isolated KSHV polypeptide comprising virion assembly protein encoded by ORF 45.

15 This invention provides the isolated KSHV polypeptide comprising glycoprotein L encoded by ORF 47.

20 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 48.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 49.

25 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF K8.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 52.

30 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 53.

35 This invention provides the isolated KSHV polypeptide comprising dUTPase encoded by ORF 54.

This invention provides the isolated KSHV polypeptide comprising major capsid protein encoded by ORF 25.

5 This invention provides the isolated KSHV polypeptide comprising capsid protein II encoded by ORF 26.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 27.

10 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 28.

This invention provides the isolated KSHV polypeptide comprising packaging protein II encoded by ORF 29b.

15 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 30.

20 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 31.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 32.

25 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 33.

This invention provides the isolated KSHV polypeptide comprising packaging protein I encoded by ORF 29a.

30 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 34.

35 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 35.

This invention provides the isolated KSHV polypeptide comprising tegument protein III encoded by ORF 64.

5 This invention provides the isolated KSHV polypeptide comprising capsid protein IV encoded by ORF 65.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 66.

10 This invention provides the isolated KSHV polypeptide comprising tegument protein IV encoded by ORF 67.

This invention provides the isolated KSHV polypeptide comprising glycoprotein encoded by ORF 68.

15 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 69.

20 This invention provides the isolated KSHV polypeptide comprising Kaposin encoded by ORF K12.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF K13.

25 This invention provides the isolated KSHV polypeptide comprising cyclin D encoded by ORF 72.

30 This invention provides the isolated KSHV polypeptide comprising immediate-early protein (IEP) encoded by ORF 73.

This invention provides the isolated KSHV polypeptide comprising OX-1 encoded by ORF K14.

35 This invention provides the isolated KSHV polypeptide comprising G-protein coupled receptor encoded by ORF 74.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF 55.

5 This invention provides the isolated KSHV polypeptide comprising DNA replication protein I encoded by ORF 56.

10 This invention provides the isolated KSHV polypeptide comprising immediate early protein II (IEP-II) encoded by ORF 57.

15 This invention provides the isolated KSHV polypeptide comprising viral interferon regulatory factor I (VIRF1; ICSEP) encoded by ORF K9. In one embodiment, VIRF1 is a transforming polypeptide.

This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF K10.

20 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF K11.

25 This invention provides the isolated KSHV polypeptide comprising phosphoprotein encoded by ORF 58.

This invention provides the isolated KSHV polypeptide comprising DNA replication protein II encoded by ORF 59.

30 This invention provides the isolated KSHV polypeptide comprising assembly/DNA maturation protein encoded by ORF 62.

35 This invention provides the isolated KSHV polypeptide comprising tegument protein II encoded by ORF 63.



This invention provides an antibody which specifically binds to the polypeptide encoded by the isolated nucleic acid molecule. In one embodiment the antibody is a monoclonal antibody. In another embodiment the antibody recognizes an epitope of the KSHV polypeptide. In another embodiment the antibody is a polyclonal antibody. In another embodiment the antibody recognizes more than one epitope of the KSHV polypeptide. In another embodiment the antibody is an anti-idiotypic antibody.

An antibody, polypeptide or isolated nucleic acid molecule may be labeled with a detectable marker including, but not limited to: a radioactive label, or a colorimetric, a luminescent, or a fluorescent marker, or gold. Radioactive labels include, but are not limited to:  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{32}\text{P}$ ,  $^{33}\text{P}$ ,  $^{35}\text{S}$ ,  $^{36}\text{Cl}$ ,  $^{51}\text{Cr}$ ,  $^{57}\text{Co}$ ,  $^{60}\text{Co}$ ,  $^{59}\text{Fe}$ ,  $^{90}\text{Y}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ , and  $^{186}\text{Re}$ . Fluorescent markers include, but are not limited to: fluorescein, rhodamine and auramine. Colorimetric markers include, but are not limited to: biotin, and digoxigenin. Methods of producing the polyclonal or monoclonal antibody are known to those of ordinary skill in the art.

Further, the antibody, polypeptide or nucleic acid molecule may be detected by a second antibody which may be linked to an enzyme, such as alkaline phosphatase or horseradish peroxidase. Other enzymes which may be employed are well known to one of ordinary skill in the art.

This invention provides a method of producing a polypeptide encoded by the isolated nucleic acid molecule, which comprises growing a host-vector system under suitable conditions permitting production of the polypeptide and recovering the polypeptide so

This invention provides the isolated KSHV polypeptide comprising tegument protein/FGARAT encoded by ORF 73.

5 This invention provides the isolated KSHV polypeptide comprising the protein encoded by ORF K13.

This invention provides the isolated KSHV polypeptide comprising viral interferon regulatory factor 2 (vIRF2) encoded by nucleotides 88,910-88,410.

10 This invention provides the isolated KSHV polypeptide comprising viral interferon regulatory factor 3 (vIRF3) encoded by nucleotides 90,541-89,600.

15 This invention provides the isolated KSHV polypeptide comprising viral interferon regulatory factor 4 (vIRF4) encoded by nucleotides 94,127-93,636.

20 This invention provides the isolated KSHV polypeptide comprising a precursor of secreted glycoprotein X (gX) encoded by nucleotides 90,173-90,643.

25 This invention provides the isolated KSHV polypeptide comprising protein T1.1 (nut-1) encoded by nucleotides 26,661-29,741.

30 Further, the isolated polypeptide may be linked to a second polypeptide to form a fusion protein by linking the isolated nucleic acid molecule to a second nucleic acid molecule and expression in a suitable host cell. In one embodiment the second nucleic acid molecule encodes beta-galactosidase. Other nucleic acid molecules which are used to form a fusion protein are  
35 known to those skilled in the art.

Polyclonal antibodies against the polypeptide may be produced by immunizing animals using a selected KSHV polypeptide. Monoclonal antibodies are prepared using hybridoma technology by fusing antibody producing B cells from immunized animals with myeloma cells and selecting the resulting hybridoma cell line producing the desired antibody, as described further below.

produced. Suitable host cells include bacteria, yeast, filamentous fungal, plant, insect and mammalian cells. Host-vector systems for producing and recovering a polypeptide are well known to those skilled in the art and include, but are not limited to, E. coli and pMAL (New England Biolabs), the Sf9 insect cell-baculovirus expression system, and mammalian cells (such as HeLa, COS, NIH 3T3 and HEK293) transfected with a mammalian expression vector by Lipofectin (Gibco-BRL) or calcium phosphate precipitation or other methods to achieve vector entry into the cell. Those of skill in the art are knowledgeable in the numerous expression systems available for expression of KSHV polypeptide.

This invention provides a method to select specific regions on the polypeptide encoded by the isolated nucleic acid molecule of the DNA virus to generate antibodies. Amino acid sequences may be analyzed by methods well known to those skilled in the art to determine whether they produce hydrophobic or hydrophilic regions in the polypeptides which they build. In the case of a cell membrane polypeptide, hydrophobic regions are well known to form the part of the polypeptide that is inserted into the lipid bilayer of the cell membrane, while hydrophilic regions are located on the cell surface, in an aqueous environment. Usually, the hydrophilic regions will be more immunogenic than the hydrophobic regions. Therefore the hydrophilic amino acid sequences may be selected and used to generate antibodies specific to polypeptide encoded by the isolated nucleic acid molecule encoding the DNA virus. The selected peptides may be prepared using commercially available machines. As an alternative, nucleic acid may be cloned and expressed and the resulting polypeptide recovered and used as an immunogen.

found in the following U.S. Pat. Nos. 4,376,110 (David et al.) or 4,098,876 (Piasio).

A. Assays for KSHV Polypeptide Antigens

5 One can use immunoassays to detect the virus, its components, or antibodies thereto. A general overview of the applicable technology is in Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publication, New York.

10 In one embodiment, antibodies to KSHV polypeptide antigens can be used. In brief, to produce antibodies, the polypeptide being targeted is expressed and purified. The product is injected into a mammal capable of producing antibodies. Either polyclonal or monoclonal antibodies (including recombinant antibodies) specific for the gene product can be used in various immunoassays. Such assays include competitive immunoassays, radioimmunoassays, Western blots, ELISA, indirect immunofluorescent assays and the like. For competitive immunoassays, see Harlow and Lane at pages 567-573 and 584-589.

25 Monoclonal antibodies or recombinant antibodies may be obtained by techniques familiar to those skilled in the art. Briefly, spleen cells or other lymphocytes from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (see, Kohler and Milstein, 1976, *Eur. J. Immunol.* 6, 511-519). Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies

30

35

## II. Immunoassays

The antibodies raised against KSHV polypeptide antigens may be detectably labeled, utilizing  
5 conventional labelling techniques well-known to the art, as described above.

In addition, enzymes may be used as labels. Suitable  
enzymes include alkaline phosphatase, beta-  
10 galactosidase, glucose-6-phosphate dehydrogenase,  
maleate dehydrogenase and peroxidase. Two principal  
types of enzyme immunoassay are the enzyme-linked  
immunosorbent assay (ELISA), and the homogeneous  
enzyme immunoassay, also known as enzyme-multiplied  
15 immunoassay (EMIT, Syva Corporation, Palo Alto, CA).  
In the ELISA system, separation may be achieved, for  
example, by the use of antibodies coupled to a solid  
phase. The EMIT system depends on deactivation of the  
enzyme in the tracer-antibody complex; activity is  
20 thus measured without the need for a separation step.

Additionally, chemiluminescent compounds may be used  
as labels. Typical chemiluminescent compounds include  
luminol, isoluminol, aromatic acridinium esters,  
25 imidazoles, acridinium salts, and oxalate esters.  
Similarly, bioluminescent compounds may be utilized  
for labelling, the bioluminescent compounds including  
luciferin, luciferase, and aequorin.

30 A description of a radioimmunoassay (RIA) may be found  
in: *Laboratory Techniques in Biochemistry and  
Molecular Biology* (1978) North Holland Publishing  
Company, New York, with particular reference to the  
chapter entitled "An Introduction to Radioimmune Assay  
and Related Techniques" by T. Chard. A description of  
35 general immunometric assays of various types can be

The protein is then purified by a combination of cell lysis (e.g., sonication) and affinity chromatography. For fusion products, subsequent digestion of the fusion protein with an appropriate proteolytic enzyme releases the desired peptide.

The polypeptides may be purified to substantial purity by standard techniques well known in the art, including selective precipitation with such substances as ammonium sulfate, column chromatography, immunopurification methods, and others. See, for instance, Scopes, 1982, Protein Purification: Principles and Practice, Springer-Verlag, New York.

B. Assays for Antibodies Specifically Binding To KSHV Polypeptides

Antibodies reactive with polypeptide antigens of KSHV can also be measured by a variety of immunoassay methods that are similar to the procedures described above for measurement of antigens. For a review of immunological and immunoassay procedures applicable to the measurement of antibodies by immunoassay techniques, see Basic and Clinical Immunology, 7th Edition, Stites and Terr, Eds., and Harlow and Lane, 1988, Antibodies, A Laboratory Manual, Cold Spring Harbor, New York.

In brief, immunoassays to measure antibodies reactive with polypeptide antigens of KSHV can be either competitive or noncompetitive binding assays. In competitive binding assays, the sample analyte competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is a purified recombinant human herpesvirus polypeptide produced as described above. Other sources of human herpesvirus

produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Newer techniques using recombinant phage antibody expression systems can also be used to generate monoclonal antibodies. See, for example: McCafferty et al. (1990) *Nature* 348, 552; Hoogenboom et al. (1991) *Nuc. Acids Res.* 19, 4133; and Marks et al. (1991) *J. Mol Biol.* 222, 581-597.

Methods for characterizing naturally processed peptides bound to MHC (major histocompatibility complex) I molecules can be used. See Falk et al., 1991, *Nature* 351, 290 and PCT publication No. WO 92/21033 published November 26, 1992. Typically, these methods involve isolation of MHC class I molecules by immunoprecipitation or affinity chromatography from an appropriate cell or cell line. Other methods involve direct amino acid sequencing of the more abundant peptides in various HPLC fractions by known automatic sequencing of peptides eluted from Class I molecules of the B cell type (Gardetzkey et al., 1991, *Nature* 353, 326), and of the human MHC class I molecule, HLA-A2.1 type by mass spectrometry (Hunt et al., 1991, *Eur. J. Immunol.* 21, 2963-2970). See also, Rötzschke and Falk, 1991, *Immunol. Today* 12, 447, for a general review of the characterization of naturally processed peptides in MHC class I. Further, Marloes et al., 1991, *Eur. J. Immunol.* 21, 2963-2970, describe how class I binding motifs can be applied to the identification of potential viral immunogenic peptides in vitro.

The polypeptides described herein produced by recombinant technology may be purified by standard techniques well known to those of skill in the art. Recombinantly produced viral polypeptides can be directly expressed or expressed as a fusion protein.



## IIA. Vector, Cell Line and Transgenic Mammal

5 This invention provides a replicable vector containing the isolated nucleic acid molecule encoding a KSHV polypeptide. The vector includes, but is not limited to: a plasmid, cosmid,  $\lambda$  phage or yeast artificial chromosome (YAC) which contains the isolated nucleic acid molecule.

10 To obtain the vector, for example, insert and vector DNA can both be exposed to a restriction enzyme to create complementary ends on both molecules which base pair with each other and are then ligated together with DNA ligase. Alternatively, linkers can be  
15 ligated to the insert DNA which correspond to a restriction site in the vector DNA, which is then digested with the restriction enzyme which cuts at that site. Other means are available and well-known to those skilled in the art.

20 This invention provides a host cell containing the vector. Suitable host cells include, but are not limited to, bacteria (such as *E. coli*), yeast, fungi, plant, insect and mammalian cells. Suitable animal  
25 cells include, but are not limited to Vero cells, HeLa cells, Cos cells, CV1 cells and various primary mammalian cells.

30 This invention provides a transgenic nonhuman mammal which comprises the isolated nucleic acid molecule introduced into the mammal at an embryonic stage. Methods of producing a transgenic nonhuman mammal are known to those skilled in the art.

35

polypeptides, including isolated or partially purified naturally occurring polypeptide, may also be used.

5 Noncompetitive assays are typically sandwich assays, in which the sample analyte is bound between two analyte-specific binding reagents. One of the binding agents is used as a capture agent and is bound to a solid surface. The second binding agent is labeled and is used to measure or detect the resultant complex  
10 by visual or instrument means. A number of combinations of capture agent and labeled binding agent can be used. A variety of different immunoassay formats, separation techniques and labels can also be used similar to those described above for the  
15 measurement of KSHV polypeptide antigens.

Hemagglutination Inhibition (HI) and Complement Fixation (CF) are two laboratory tests that can be used to detect infection with human herpesvirus by  
20 testing for the presence of antibodies against the virus or antigens of the virus.

Serological methods can also be useful when one wishes to detect antibody to a specific viral variant. For  
25 example, one may wish to see how well a vaccine recipient has responded to a new preparation by assay of patient sera.

In one embodiment the nucleic acid molecule from the tumor lesion is amplified before step (b). In another embodiment the polymerase chain reaction (PCR) is employed to amplify the nucleic acid molecule.  
5 Methods of amplifying nucleic acid molecules are known to those skilled in the art.

A person of ordinary skill in the art will be able to obtain appropriate nucleic acid sample for diagnosing Kaposi's sarcoma in the subject. The DNA sample  
10 obtained by the above described method may be cleaved by restriction enzyme before analysis, a technique well-known in the art.

In the above described methods, a size fractionation may be employed which is effected by a polyacrylamide gel. In one embodiment, the size fractionation is effected by an agarose gel. Further, transferring the nucleic acid fragments into a solid matrix may be  
15 employed before a hybridization step. One example of such solid matrix is nitrocellulose paper.  
20

This invention provides a method of detecting expression of a KSHV gene in a cell which comprises obtaining mRNA from the cell, contacting the mRNA with a labeled nucleic acid molecule of KSHV under hybridizing conditions, determining the presence of mRNA hybridized to the molecule, thereby detecting expression of the KSHV gene. In one embodiment cDNA  
25 is prepared from the mRNA obtained from the cell and used to detect KSHV expression.  
30

Accepted means for conducting hybridization assays are known and general overviews of the technology can be had from a review of: Nucleic Acid Hybridization: A  
35 Practical Approach (1985) Hames and Higgins, Eds., IRL Press; Hybridization of Nucleic Acids Immobilized on

### III. Diagnostic Assays for KS

This invention embraces diagnostic test kits for detecting the presence of KSHV in biological samples, such as skin samples or samples of other affected tissue, comprising a container containing a nucleic acid sequence specific for a KSHV polypeptide and instructional material for performing the test. A container containing nucleic acid primers to any one of such sequences is optionally included.

This invention further embraces diagnostic test kits for detecting the presence of KSHV in biological samples, such as serum or solid tissue samples, comprising a container containing antibodies to a KSHV polypeptide, and instructional material for performing the test. Alternatively, inactivated viral particles or polypeptides derived from the human herpesvirus may be used in a diagnostic test kit to detect antibodies specific for a KSHV polypeptide.

#### A. Nucleic Acid Assays

This invention provides a method of diagnosing Kaposi's sarcoma in a subject which comprises: (a) obtaining a nucleic acid molecule from a tumor lesion or a suitable bodily fluid of the subject; (b) contacting the nucleic acid molecule with a labeled nucleic acid molecule of at least 15 nucleotides capable of specifically hybridizing with the isolated nucleic acid molecule of KSHV under hybridizing conditions; and (c) determining the presence of the nucleic acid molecule hybridized, the presence of which is indicative of Kaposi's sarcoma in the subject, thereby diagnosing Kaposi's sarcoma in the subject.

Vol. 152, (1987) Academic Press, New York; or  
*Hybridization with Nucleic Acid Probes*, pp. 495-524,  
(1993) Elsevier, Amsterdam.

5 Usually, at least a part of the probe will have  
considerable sequence identity with the target nucleic  
acid. Although the extent of the sequence identity  
required for specific hybridization will depend on the  
length of the probe and the hybridization conditions,  
10 the probe will usually have at least 70% identity to  
the target nucleic acid, more usually at least 80%  
identity, still more usually at least 90% identity and  
most usually at least 95% or 100% identity.

15 The following stringent hybridization and washing  
conditions will be adequate to distinguish a specific  
probe (e.g., a fluorescently labeled nucleic acid  
probe) from a probe that is not specific: incubation  
of the probe with the sample for 12 hours at 37°C in  
20 a solution containing denatured probe, 50% formamide,  
2X SSC, and 0.1% (w/v) dextran sulfate, followed by  
washing in 1X SSC at 70°C for 5 minutes; 2X SSC at  
37°C for 5 minutes; 0.2X SSC at room temperature for  
5 minutes, and H<sub>2</sub>O at room temperature for 5 minutes.  
25 Those of skill are aware that it will often be  
advantageous in nucleic acid hybridizations (i.e., in  
situ, Southern, or Northern) to include detergents  
(e.g., sodium dodecyl sulfate), chelating agents  
(e.g., EDTA) or other reagents (e.g., buffers,  
30 Denhardt's solution, dextran sulfate) in the  
hybridization or wash solutions. To evaluate  
specificity, probes can be tested on host cells  
containing KSHV and compared with the results from  
cells containing non-KSHV virus.

35

It will be apparent to those of ordinary skill in the  
art that a convenient method for determining whether

Solid Supports, Meinkoth and Wahl; *Analytical Biochemistry* (1984) 236, 267-284 and Innis et al., *PCR Protocols* (1990) Academic Press, San Diego.

5 Target-specific probes may be used in the nucleic acid hybridization diagnostic assays for KS. The probes are specific for or complementary to the target of interest. For precise allelic differentiations, the probes should be about 14 nucleotides long and  
10 preferably about 20-30 nucleotides. For more general detection of KSHV, nucleic acid probes are about 50 to 1000 nucleotides, most preferably about 200 to 400 nucleotides.

15 A specific nucleic acid probe can be RNA, DNA, oligonucleotide, or their analogs. The probes may be single or double stranded nucleic acid molecules. The probes of the invention may be synthesized enzymatically, using methods well known in the art  
20 (e.g., nick translation, primer extension, reverse transcription, the polymerase chain reaction, and others) or chemically (e.g., by methods described by Beaucage and Carruthers or Matteucci et al., supra).

25 The probe must be of sufficient length to be able to form a stable duplex with its target nucleic acid in the sample, i.e., at least about 14 nucleotides, and may be longer (e.g., at least about 50 or 100 bases in length). Often the probe will be more than about 100  
30 bases in length. For example, when probe is prepared by nick-translation of DNA in the presence of labeled nucleotides the average probe length may be about 100-600 bases.

35 For discussions of nucleic acid probe design and annealing conditions see, for example, Ausubel et al., supra; Berger and Kimmel, Eds., *Methods in Enzymology*

An alternative means for determining the presence of the human herpesvirus is in situ hybridization, or more recently, in situ polymerase chain reaction. In situ PCR is described in Neuvo et al. (1993) Intracellular localization of PCR-amplified hepatitis C DNA, in *American Journal of Surgical Pathology* 17(7), 683-690; Bagasra et al. (1992) Detection of HIV-1 provirus in mononuclear cells by in situ PCR, in *New England Journal of Medicine* 326(21), 1385-1391; and Heniford et al. (1993) Variation in cellular EGF receptor mRNA expression demonstrated by in situ reverse transcriptase polymerase chain reaction, in *Nucleic Acids Research* 21, 3159-3166. In situ hybridization assays are well known and are generally described in *Methods Enzymol.* Vol. 152, (1987) Berger and Kimmel, Eds., Academic Press, New York. In an in situ hybridization, cells are fixed to a solid support, typically a glass slide. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of target-specific probes that are labeled. The probes are preferably labeled with radioisotopes or fluorescent reporters.

The above-described probes are also useful for in situ hybridization or in order to locate tissues which express the gene, or for other hybridization assays for the presence of the gene or its mRNA in various biological tissues. In situ hybridization is a sensitive localization method which is not dependent on expression of polypeptide antigens or native versus denatured conditions.

Synthetic oligonucleotide (oligo) probes and riboprobes made from KSHV phagemids or plasmids are also provided. Successful hybridization conditions in tissue sections is readily transferrable from one probe to another. Commercially-synthesized

a probe is specific for a KSHV nucleic acid molecule utilizes a Southern blot (or Dot blot) using DNA prepared from the virus. Briefly, to identify a target-specific probe, DNA is isolated from the virus. Test DNA, either viral or cellular, is transferred to a solid (e.g., charged nylon) matrix. The probes are labeled by conventional methods. Following denaturation and/or prehybridization steps known in the art, the probe is hybridized to the immobilized DNAs under stringent conditions, such as defined above.

It is further appreciated that in determining probe specificity and in utilizing the method of this invention to detect KSHV, a certain amount of background signal is typical and can easily be distinguished by one of skill from a specific signal. Two-fold signal over background is acceptable.

A preferred method for detecting the KSHV polypeptide is the use of PCR and/or dot blot hybridization. Other methods to test for the presence or absence of KSHV for detection or prognosis, or risk assessment for KS includes Southern transfers, solution hybridization or non-radioactive detection systems, all of which are well known to those of skill in the art. Hybridization is carried out using probes. Visualization of the hybridized portions allows the qualitative determination of the presence or absence of the causal agent.

Similarly, a Northern transfer or reverse transcriptase PCR may be used for the detection of KSHV messenger RNA in a sample. These procedures are also well known in the art. See Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2nd ed.), Cold Spring Harbor Laboratory, Vols. 1-3.



Alternative immunohistochemical protocols may be employed which are well known to those skilled in the art.

5           B.   Immunologic Assays

This invention provides a method of diagnosing Kaposi's sarcoma in a subject, which comprises (a) obtaining a suitable bodily fluid sample from the  
10       subject, (b) contacting the suitable bodily fluid of the subject to a support having already bound thereto an antibody recognizing the KSHV polypeptide, so as to bind the antibody to a specific KSHV polypeptide antigen, (c) removing unbound bodily fluid from the  
15       support, and (d) determining the level of the antibody bound by the antigen, thereby diagnosing Kaposi's sarcoma.

This invention provides a method of diagnosing Kaposi's sarcoma in a subject, which comprises (a)  
20       obtaining a suitable bodily fluid sample from the subject, (b) contacting the suitable bodily fluid of the subject to a support having already bound thereto the KSHV polypeptide antigen, so as to bind the  
25       antigen to a specific Kaposi's sarcoma antibody, (c) removing unbound bodily fluid from the support, and (d) determining the level of the antigen bound by the Kaposi's sarcoma antibody, thereby diagnosing Kaposi's sarcoma.

30       The suitable bodily fluid sample is any bodily fluid sample which would contain Kaposi's sarcoma antibody, antigen or fragments thereof. A suitable bodily fluid includes, but is not limited to: serum, plasma, cerebrospinal fluid, lymphocytes, urine, transudates,  
35       or exudates. In the preferred embodiment, the suitable bodily fluid sample is serum or plasma. In

oligonucleotide probes are prepared using the nucleotide sequence of the identified gene. These probes are chosen for length (45-65 mers), high G-C content (50-70%) and are screened for uniqueness against other viral sequences in GenBank.

Oligos are 3'-end-labeled with [ $\alpha$ - $^{35}$ S]dATP to specific activities in the range of  $1 \times 10^{10}$  dpm/ $\mu$ g using terminal deoxynucleotidyl transferase. Unincorporated labeled nucleotides are removed from the oligo probe by centrifugation through a Sephadex G-25 column or by elution from a Waters Sep Pak C-18 column.

KS tissue embedded in OCT compound and snap frozen in freezing isopentane cooled with dry ice is cut at 6  $\mu$ m intervals and thawed onto 3-aminopropyltriethoxysilane treated slides and allowed to air dry. The slides are then fixed in 4% freshly prepared paraformaldehyde and rinsed in water. Formalin-fixed, paraffin embedded KS tissues cut at 6  $\mu$ m and baked onto glass slides can also be used. These sections are then deparaffinized in xylenes and rehydrated through graded alcohols. Prehybridization in 20mM Tris pH 7.5, 0.02% Denhardt's solution, 10% dextran sulfate for 30 min at 37°C is followed by hybridization overnight in a solution of 50% formamide (v/v), 10% dextran sulfate (w/v), 20mM sodium phosphate (pH 7.4), 3X SSC, 1X Denhardt's solution, 100  $\mu$ g/ml salmon sperm DNA, 125  $\mu$ g/ml yeast tRNA and the oligo probe ( $10^6$  cpm/ml) at 42°C overnight. The slides are washed twice with 3X SSC and twice with 1X SSC for 15 minutes each at room temperature and visualized by autoradiography. Briefly, sections are dehydrated through graded alcohols containing 0.3M ammonium acetate, and air dried. The slides are dipped in Kodak NTB2 emulsion, exposed for days to weeks, developed, and counterstained with hematoxylin and eosin (H&E).

See Immunoassays above for more details on the immunoreagents of the invention for use in diagnostic assays for KS.

5      IV.    Treatment of Human Herpesvirus-Induced KS

10      This invention provides a method for treating a subject with Kaposi's sarcoma (KS) comprising administering to the subject having KS a pharmaceutically effective amount of an antiviral agent in a pharmaceutically acceptable carrier, wherein the agent is effective to treat the subject with KSHV.

15      Further, this invention provides a method of prophylaxis or treatment for Kaposi's sarcoma (KS) by administering to a patient at risk for KS, an antibody that binds to KSHV in a pharmaceutically acceptable carrier.

20      This invention provides a method of treating a subject with Kaposi's sarcoma comprising administering to the subject an effective amount of an antisense molecule capable of hybridizing to the isolated DNA molecule of KSHV under conditions such that the antisense molecule selectively enters a KS tumor cell of the  
25      subject, so as to treat the subject.

addition, the sample may be cells from bone marrow, or a supernatant from a cell culture. Methods of obtaining a suitable bodily fluid sample from a subject are known to those skilled in the art. Methods of determining the level of antibody or antigen include, but are not limited to: ELISA, IFA, and Western blotting. Other methods are known to those skilled in the art. Further, a subject infected with KSHV may be diagnosed as infected with the above-described methods.

The detection of KSHV and the detection of virus-associated KS are essentially identical processes. The basic principle is to detect the virus using specific ligands that bind to the virus but not to other polypeptides or nucleic acids in a normal human cell or its environs. The ligands can be nucleic acid molecules, polypeptides or antibodies. The ligands can be naturally-occurring or genetically or physically modified, such as nucleic acids with non-natural nucleotide bases or antibody derivatives, i.e., Fab or chimeric antibodies. Serological tests for detection of antibodies to the virus present in subject sera may also be performed by using the KSHV polypeptide as an antigen, as described herein.

Samples can be taken from patients with KS or from patients at risk for KS, such as AIDS patients. Typically the samples are taken from blood (cells, serum and/or plasma) or from solid tissue samples such as skin lesions. The most accurate diagnosis for KS will occur if elevated titers of the virus are detected in the blood or in involved lesions. KS may also be indicated if antibodies to the virus are detected and if other diagnostic factors for KS are present.

promoting inhibitory mechanisms of the cells, such as promoting RNA degradation. Inhibitory nucleic acid methods therefore encompass a number of different approaches to altering expression of herpesvirus genes. These different types of inhibitory nucleic acid technology are described in Helene and Toulme (1990) *Biochim. Biophys. Acta.* 1049, 99-125, which is referred to hereinafter as "Helene and Toulme."

In brief, inhibitory nucleic acid therapy approaches can be classified into those that target DNA sequences, those that target RNA sequences (including pre-mRNA and mRNA), those that target proteins (sense strand approaches), and those that cause cleavage or chemical modification of the target nucleic acids.

Approaches targeting DNA fall into several categories. Nucleic acids can be designed to bind to the major groove of the duplex DNA to form a triple helical or "triplex" structure. Alternatively, inhibitory nucleic acids are designed to bind to regions of single stranded DNA resulting from the opening of the duplex DNA during replication or transcription.

More commonly, inhibitory nucleic acids are designed to bind to mRNA or mRNA precursors. Inhibitory nucleic acids are used to prevent maturation of pre-mRNA. Inhibitory nucleic acids may be designed to interfere with RNA processing, splicing or translation.

The inhibitory nucleic acids can be targeted to mRNA. In this approach, the inhibitory nucleic acids are designed to specifically block translation of the encoded protein. Using this approach, the inhibitory nucleic acid can be used to selectively suppress certain cellular functions by inhibition of

### A. Nucleic Acid Therapeutics

This invention provides an antisense molecule capable of hybridizing to the isolated nucleic acid molecule of KSHV. In one embodiment the antisense molecule is DNA. In another embodiment the antisense molecule is RNA. In another embodiment, the antisense molecule is a nucleic acid derivative (e.g., DNA or RNA with a protein backbone).

The present invention extends to the preparation of antisense nucleic acids and ribozymes that may be used to interfere with the expression of a polypeptide either by masking the mRNA with an antisense nucleic acid or cleaving it with a ribozyme, respectively.

This invention provides inhibitory nucleic acid therapeutics which can inhibit the activity of herpesviruses in patients with KS by binding to the isolated nucleic acid molecule of KSHV. Inhibitory nucleic acids may be single-stranded nucleic acids, which can specifically bind to a complementary nucleic acid sequence. By binding to the appropriate target sequence, an RNA-RNA, a DNA-DNA, or RNA-DNA duplex or triplex is formed. These nucleic acids are often termed "antisense" because they are usually complementary to the sense or coding strand of the gene, although recently approaches for use of "sense" nucleic acids have also been developed. The term "inhibitory nucleic acids" as used herein, refers to both "sense" and "antisense" nucleic acids.

By binding to the target nucleic acid, the inhibitory nucleic acid can inhibit the function of the target nucleic acid. This could, for example, be a result of blocking DNA transcription, processing or poly(A) addition to mRNA, DNA replication, translation, or

The targeting of inhibitory nucleic acids to specific cells of the immune system by conjugation with targeting moieties binding receptors on the surface of these cells can be used for all of the above forms of inhibitory nucleic acid therapy. This invention encompasses all of the forms of inhibitory nucleic acid therapy as described above and as described in Helene and Touime.

10 An example of an antiherpes virus inhibitory nucleic acid is ISIS 2922 (ISIS Pharmaceuticals) which has activity against CMV (see *Biotechnology News* 14:5).

15 A problem associated with inhibitory nucleic acid therapy is the effective delivery of the inhibitory nucleic acid to the target cell in vivo and the subsequent internalization of the inhibitory nucleic acid by that cell. This can be accomplished by linking the inhibitory nucleic acid to a targeting moiety to form a conjugate that binds to a specific receptor on the surface of the target infected cell, and which is internalized after binding.

### 25 B. Antiviral Agents

25 The use of combinations of antiviral drugs and sequential treatments are useful for treatment of herpesvirus infections and will also be useful for the treatment of herpesvirus-induced KS. For example, Snoeck et al. (1992) *Eur. J. Clin. Micro. Infect. Dis.* 11, 1144-1155, found additive or synergistic effects against CMV when combining antiherpes drugs (e.g., combinations of zidovudine [3'-azido-3'-deoxythymidine, AZT] with HPMPC, ganciclovir, 30 foscarnet or acyclovir or of HPMPC with other antivirals). Similarly, in treatment of cytomegalovirus retinitis, induction with ganciclovir

translation of mRNA encoding critical proteins. For example, an inhibitory nucleic acid complementary to regions of c-myc mRNA inhibits c-myc protein expression in a human promyelocytic leukemia cell line, HL60, which overexpresses the c-myc proto-oncogene. See Wickstrom et al. (1988) PNAS 85, 1028-1032 and Harel-Bellan et al. (1988) Exp. Med. 168, 2309-2318. As described in Helene and Toulme, inhibitory nucleic acids targeting mRNA have been shown to work by several different mechanisms to inhibit translation of the encoded protein(s).

The inhibitory nucleic acids introduced into the cell can also encompass the "sense" strand of the gene or mRNA to trap or compete for the enzymes or binding proteins involved in mRNA translation, as described in Helene and Toulme.

Lastly, the inhibitory nucleic acids can be used to induce chemical inactivation or cleavage of the target genes or mRNA. Chemical inactivation can occur by the induction of crosslinks between the inhibitory nucleic acid and the target nucleic acid within the cell. Other chemical modifications of the target nucleic acids induced by appropriately derivatized inhibitory nucleic acids may also be used.

Cleavage, and therefore inactivation, of the target nucleic acids may be effected by attaching a substituent to the inhibitory nucleic acid which can be activated to induce cleavage reactions. The substituent can be one that affects either chemical, or enzymatic cleavage. Alternatively, cleavage can be induced by the use of ribozymes or catalytic RNA. In this approach, the inhibitory nucleic acids would comprise either naturally occurring RNA (ribozymes) or synthetic nucleic acids with catalytic activity.



Infectious Disease Ch.35, 289, W.B. Saunders, Philadelphia, Pennsylvania) and the like. Immunological therapy will also be effective in many cases to manage and alleviate symptoms caused by the disease agents described here. Antiviral agents include agents or compositions that directly bind to viral products and interfere with disease progress; and, excludes agents that do not impact directly on viral multiplication or viral titer. Antiviral agents do not include immunoregulatory agents that do not directly affect viral titer or bind to viral products. Antiviral agents are effective if they inactivate the virus, otherwise inhibit its infectivity or multiplication, or alleviate the symptoms of KS.

The antiherpesvirus agents that will be useful for treating virus-induced KS can be grouped into broad classes based on their presumed modes of action. These classes include agents that act (1) by inhibition of viral DNA polymerase, (2) by targeting other viral enzymes and proteins, (3) by miscellaneous or incompletely understood mechanisms, or (4) by binding a target nucleic acid (i.e., inhibitory nucleic acid therapeutics, supra). Antiviral agents may also be used in combination (i.e., together or sequentially) to achieve synergistic or additive effects or other benefits.

Although it is convenient to group antiviral agents by their supposed mechanism of action, the applicants do not intend to be bound by any particular mechanism of antiviral action. Moreover, it will be understood by those of skill that an agent may act on more than one target in a virus or virus-infected cell or through more than one mechanism.

#### 1) Inhibitors of DNA Polymerase

5 followed by maintenance with foscarnet has been suggested as a way to maximize efficacy while minimizing the adverse side effects of either treatment alone. An anti-herpetic composition that contains acyclovir and, e.g., 2-acetylpyridine-5-(2-pyridylamino)thiocarbonyl)-thiocarbonchydrazone is described in U.S. Pat. 5,175,165 (assigned to Burroughs Wellcome Co.). Combinations of TS-inhibitors and viral TK-inhibitors in antiherpetic medicines are disclosed in U.S. Pat. 5,137,724, assigned to Stichting Rega VZW. A synergistic inhibitory effect on EBV replication using certain ratios of combinations of HPMPC with AZT was reported by Lin et al. (1991) Antimicrob Agents Chemother 35:2440-3.

20 U.S. Patent Nos. 5,164,395 and 5,021,437 (Blumenkopf; Burroughs Wellcome) describe the use of a ribonucleotide reductase inhibitor (an acetylpyridine derivative) for treatment of herpes infections, including the use of the acetylpyridine derivative in combination with acyclovir. U.S. Patent No. 5,137,724 (Balzari et al. (1990) Mol. Pharm. 37,402-7) describes the use of thymidylate synthase inhibitors (e.g., 5-fluoro-uracil and 5-fluoro-2'-deoxyuridine) in combination with compounds having viral thymidine kinase inhibiting activity.

30 With the discovery of a disease causal agent for KS now identified, effective therapeutic or prophylactic protocols to alleviate or prevent the symptoms of herpes virus-associated KS can be formulated. Due to the viral nature of the disease, antiviral agents have application here for treatment, such as interferons, nucleoside analogues, ribavirin, amantadine, and pyrophosphate analogues of phosphonoacetic acid (foscarnet) (reviewed in Gorbach et al., 1992,

antiviral drug. The mechanism of action of certain anti-herpesvirus agents is discussed in De Clercq (1993, *Antimicrobial Chemotherapy* 33, Suppl. A, 121-132) and in other references cited *supra* and *infra*.

5

Anti-herpesvirus medications suitable for treating viral induced KS include, but are not limited to, nucleoside analogs including acyclic nucleoside phosphonate analogs (e.g., phosphonyl-methoxyalkylpurines and -pyrimidines), and cyclic nucleoside analogs. These include drugs such as: vidarabine (9- $\beta$ -D-arabinofuranosyladenine; adenine arabinoside, ara-A, Vira-A, Parke-Davis); 1- $\beta$ -D-arabinofuranosyluracil (ara-U); 1- $\beta$ -D-arabinofuranosyl-cytosine (ara-C); HPMPC [(S)-1-[3-hydroxy-2-(phosphonylmethoxy)propyl]cytosine (e.g., GS 504, Gilead Science)] and its cyclic form (cHPMPC); HPMPA [(S)-9-(3-hydroxy-2-phosphonylmethoxypropyl)adenine] and its cyclic form (cHPMPA); (S)-HPMPDAP [(S)-9-(3-hydroxy-2-phosphonylmethoxypropyl)-2,6-diaminopurine]; PMEDAP [9-(2-phosphonyl-methoxyethyl)-2,6-diaminopurine]; HOE 602 [2-amino-9-(2,3-bis(isopropoxy)-2-propoxymethyl)purine]; PMEA [9-(2-phosphonylmethoxyethyl)adenine]; bromovinyl-deoxyuridine (Burns and Sandford, 1990, *J. Infect. Dis.* 162:634-7); 1- $\beta$ -D-arabinofuranosyl-E-5-(2-bromovinyl)-uridine or -2'-deoxyuridine; BVaraU (1- $\beta$ -D-arabinofuranosyl-E-5-(2-bromovinyl)-uracil, brovavir, Bristol-Myers Squibb, Yamsa Shoyu); BVDU [(E)-5-(2-bromovinyl)-2'-deoxyuridine, brivudin, e.g., Helpin] and its carbocyclic analogue (in which the sugar moiety is replaced by a cyclopentane ring); IVDU [(E)-5-(2-iodovinyl)-2'-deoxyuridine] and its carbocyclic analogue, C-IVDU (Balzarini et al., *supra*); and 5-mercaptithio analogs of 2'-deoxyuridine (Holliday and Williams, 1992, *Antimicrob. Agents Chemother.* 36, 1935); acyclovir [9-(2-

35

Many antiherpesvirus agents in clinical use or in development today are nucleoside analogs believed to act through inhibition of viral DNA replication, especially through inhibition of viral DNA polymerase. These nucleoside analogs act as alternative substrates for the viral DNA polymerase or as competitive inhibitors of DNA polymerase substrates. Usually these agents are preferentially phosphorylated by viral thymidine kinase (TK), if one is present, and/or have higher affinity for viral DNA polymerase than for the cellular DNA polymerases, resulting in selective antiviral activity. Where a nucleoside analogue is incorporated into the viral DNA, viral activity or reproduction may be affected in a variety of ways. For example, the analogue may act as a chain terminator, cause increased lability (e.g., susceptibility to breakage) of analogue-containing DNA, and/or impair the ability of the substituted DNA to act as template for transcription or replication (see, e.g., Balzarini et al., *supra*).

It will be known to one of skill that, like many drugs, many of the agents useful for treatment of herpes virus infections are modified (i.e., "activated") by the host, host cell, or virus-infected host cell metabolic enzymes. For example, acyclovir is triphosphorylated to its active form, with the first phosphorylation being carried out by the herpes virus thymidine kinase, when present. Other examples are the reported conversion of the compound HOE 602 to ganciclovir in a three-step metabolic pathway (Winkler et al., 1990, *Antiviral Research* 14, 61-74) and the phosphorylation of ganciclovir to its active form by, e.g., a CMV nucleotide kinase. It will be apparent to one of skill that the specific metabolic capabilities of a virus can affect the sensitivity of that virus to specific drugs, and is one factor in the choice of an

Human Retroviruses 9, 307-314) and are additional nucleoside analogs that may be used to treat KS. An exemplary protocol for these agents is an intravenous injection of about 0.35 mg/meter<sup>2</sup> (0.7 mg/kg) once weekly or every other week for at least two doses, preferably up to about four to eight weeks.

Acyclovir and ganciclovir are of interest because of their accepted use in clinical settings. Acyclovir, an acyclic analogue of guanine, is phosphorylated by a herpesvirus thymidine kinase and undergoes further phosphorylation to be incorporated as a chain terminator by the viral DNA polymerase during viral replication. It has therapeutic activity against a broad range of herpesviruses, Herpes simplex Types 1 and 2, Varicella-Zoster, Cytomegalovirus, and Epstein-Barr Virus, and is used to treat disease such as herpes encephalitis, neonatal herpesvirus infections, chickenpox in immunocompromised hosts, herpes zoster recurrences, CMV retinitis, EBV infections, chronic fatigue syndrome, and hairy leukoplakia in AIDS patients. Exemplary intravenous dosages or oral dosages are 250 mg/kg/m<sup>2</sup> body surface area, every 8 hours for 7 days, or maintenance doses of 200-400 mg IV or orally twice a day to suppress recurrence. Ganciclovir has been shown to be more active than acyclovir against some herpesviruses. See, e.g., Oren and Soble, 1991, *Clinical Infectious Diseases* 14, 741-6. Treatment protocols for ganciclovir are 5 mg/kg twice a day IV or 2.5 mg/kg three times a day for 10-14 days. Maintenance doses are 5-6 mg/kg for 5-7 days.

Also of interest is HPMPC. HPMPC is reported to be more active than either acyclovir or ganciclovir in the chemotherapy and prophylaxis of various HSV-1,

hydroxyethoxy)methyl]guanine; e.g., Zovirax (Burroughs Wellcome)); penciclovir (9-[4-hydroxy-3-(hydroxymethyl)butyl]-guanine); ganciclovir (9-[1,3-dihydroxy-2 propoxymethyl]-guanine) e.g., Cymevene, Cytovene (Syntex), DHPG (Stals et al., 1993, Antimicrobial Agents Chemother. 37, 218-223; isopropylether derivatives of ganciclovir (see, e.g., Winkelmann et al., 1988, Drug Res. 38, 1545-1548); cygalovir; famciclovir [2-amino-9-(4-acetoxy-3-(acetoxymethyl)but-1-yl)purine (Smithkline Beecham)]; valacyclovir (Burroughs Wellcome); dasciclovir [(2-amino-9-(2-ethoxymethyl)purine)] and 2-amino-9-(2-hydroxyethoxymethyl)-9H-purine, prodrugs of acyclovir]; CDG (carbocyclic 2'-deoxyguanosine); and purine nucleosides with the pentafuranosyl ring replaced by a cyclobutane ring (e.g., cyclobut-A [(--)-9-[1 $\beta$ , 2 $\alpha$ , 3 $\beta$ ]-2,3-bis(hydroxymethyl)-1-cyclobutyl]adenine], cyclobut-G [(--)-9-[1 $\beta$ , 2 $\alpha$ , 3 $\beta$ ]-2,3-bis(hydroxymethyl)-1-cyclobutyl]guanine], BHCG [(R)-9-[1 $\alpha$ , 2 $\beta$ , 3 $\alpha$ ]-2,3-bis(hydroxymethyl)cyclobutyl]guanine], and an active isomer of racemic BHCG, SQ 34,514 [(1R-1 $\alpha$ , 2 $\beta$ , 3 $\alpha$ )-2-amino-9-[2,3-bis(hydroxymethyl)cyclobutyl]-6H-purin-6-one (see, Braitman et al., 1991, Antimicrob. Agents and Chemotherapy 35, 1464-1468). Certain of these antiherpesviral agents are discussed in Gorach et al., 1992, Infectious Disease Ch.35, 289, W.B. Saunders, Philadelphia; Saunders et al., 1990, J. Acquir. Immune Defic. Syndr. 3, 571; Yamanaka et al., 1991, Mol. Pharmacol. 40, 446; and Greenspan et al., 1990, J. Acquir. Immune Defic. Syndr. 3, 571.

Triciribine and triciribine monophosphate are potent inhibitors against herpes viruses. (Ickes et al., 1994, Antiviral Research 23, Seventh International Conf. on Antiviral Research, Abstract No. 122, Supp. 1.), HIV-1 and HIV-2 (Kucera et al., 1993, AIDS Res.

MISSING UPON TIME OF PUBLICATION

HSV-2, TK- HSV, VZV or CMV infections in animal models (De Clercq, supra).

5 Nucleoside analogs such as BVaraT are potent inhibitors of HSV-1, EBV, and VZV that have greater activity than acyclovir in animal models of encephalitis. FIAU (fluoridoarabinosyl cytosine and its related fluoroethyl and iodo compounds (e.g., FEAU, FIAU) have potent selective activity against  
10 herpesviruses, and HPMPA ((S)-1-([3-hydroxy-2-phosphorylmethoxy]propyl)adenine) has been demonstrated to be more potent against HSV and CMV than acyclovir or ganciclovir and are of choice in advanced cases of KS. Cladribine (2-chlorodeoxyadenosine)  
15 is another nucleoside analogue known as a highly specific antilymphocyte agent (i.e., a immunosuppressive drug).

Other useful antiviral agents include: 5-thien-2-yl-2'-deoxyuridine derivatives, e.g., BTDU [5-(5-bromothien-2-yl)-2'-deoxyuridine] and CTDU [5-(5-chlorothien-2-yl)-2'-deoxyuridine]; and OXT-A [9-(2-deoxy-2-hydroxymethyl- $\beta$ -D-erythro-oxetanosyl)adenine] and OXT-G [9-(2-deoxy-2-hydroxymethyl- $\beta$ -D-erythro-oxetanosyl)guanine].  
20 Although OXT-G is believed to act by inhibiting viral DNA synthesis its mechanism of action has not yet been elucidated. These and other compounds are described in Andrei et al., 1992, Eur. J. Clin. Microbiol. Infect. Dis. 11, 143-51.  
30 Additional antiviral purine derivatives useful in treating herpesvirus infections are disclosed in US Pat. 5,108,994 (assigned to Beecham Group P.L.C.). 6-Methoxypurine arabinoside (ara-M; Burroughs Wellcome) is a potent inhibitor of varicella-zoster virus, and  
35 will be useful for treatment of KS.



### ii) Other Antivirals

Although applicants do not intend to be bound by a particular mechanism of antiviral action, the antiherpes-virus agents described above are believed to act through inhibition of viral DNA polymerase. However, viral replication requires not only the replication of the viral nucleic acid but also the production of viral proteins and other essential components. Accordingly, the present invention contemplates treatment of KS by the inhibition of viral proliferation by targeting viral proteins other than DNA polymerase (e.g., by inhibition of their synthesis or activity, or destruction of viral proteins after their synthesis). For example, administration of agents that inhibit a viral serine protease, e.g., such as one important in development of the viral capsid will be useful in treatment of viral induced KS.

Other viral enzyme targets include: OMP decarboxylase inhibitors (a target of, e.g., parazofurin), CTP synthetase inhibitors (targets of, e.g., cyclopentenylcytosine), IMP dehydrogenase, ribonucleotide reductase (a target of, e.g., carboxyl-containing N-alkyldipeptides as described in U.S. Patent No. 5,110,799 (Tolman et al., Merck)), thymidine kinase (a target of, e.g., 1-[2-(hydroxymethyl)cycloalkylmethyl]-5-substituted -uracils and -guanines as described in, e.g., U.S. Patent Nos. 4,863,927 and 4,782,062 (Tolman et al., Merck) as well as other enzymes. It will be apparent to one of ordinary skill in the art that there are additional viral proteins, both characterized and as yet to be discovered, that can serve as target for antiviral agents.

Kutapressin is a liver derivative available from Schwarz Parma of Milwaukee, Wisconsin in an injectable form of 25 mg/ml. The recommended dosage for herpesviruses is from 200 to 25 mg/ml per day for an average adult of 150 pounds.

Poly(I) Poly(C<sub>12</sub>U), an accepted antiviral drug known as Ampligen from HEM Pharmaceuticals of Rockville, MD has been shown to inhibit herpesviruses and is another antiviral agent suitable for treating KS. Intravenous injection is the preferred route of administration. Dosages from about 100 to 600 mg/m<sup>2</sup> are administered two to three times weekly to adults averaging 150 pounds. It is best to administer at least 200 mg/m<sup>2</sup> per week.

Other antiviral agents reported to show activity against herpes viruses (e.g., varicella zoster and herpes simplex) and will be useful for the treatment of herpesvirus-induced KS include mappicine ketone (SmithKline Beecham); Compounds A,79296 and A,73209 (Abbott) for varicella zoster, and Compound 882087 (Burroughs Wellcome) (see, The Pink Sheet 55(20) May 17, 1993).

Interferon is known inhibit replication of herpes viruses. See Oren and Soble, supra. Interferon has known toxicity problems and it is expected that second generation derivatives will soon be available that will retain interferon's antiviral properties but have reduced side affects.

It is also contemplated that herpes virus-induced KS may be treated by administering a herpesvirus reactivating agent to induce reactivation of the latent virus. Preferably the reactivation is combined

As used herein administration means a method of administering to a subject. Such methods are well known to those skilled in the art and include, but are not limited to, administration topically, parenterally, orally, intravenously, intramuscularly, subcutaneously or by aerosol. Administration of the agent may be effected continuously or intermittently such that the therapeutic agent in the patient is effective to treat a subject with Kaposi's sarcoma or a subject infected with a DNA virus associated with Kaposi's sarcoma.

The antiviral compositions for treating herpesvirus-induced KS are preferably administered to human patients via oral, intravenous or parenteral administrations and other systemic forms. Those of skill in the art will understand appropriate administration protocol for the individual compositions to be employed by the physician.

The pharmaceutical formulations or compositions of this invention may be in the dosage form of solid, semi-solid, or liquid such as, e.g., suspensions, aerosols or the like. Preferably the compositions are administered in unit dosage forms suitable for single administration of precise dosage amounts. The compositions may also include, depending on the formulation desired, pharmaceutically-acceptable, non-toxic carriers or diluents, which are defined as vehicles commonly used to formulate pharmaceutical compositions for animal or human administration. The diluent is selected so as not to affect the biological activity of the combination. Examples of such diluents are distilled water, physiological saline, Ringer's solution, dextrose solution, and Hank's solution. In addition, the pharmaceutical composition or formulation may also include other carriers,

with simultaneous or sequential administration of an anti-herpesvirus agent. Controlled reactivation over a short period of time or reactivation in the presence of an antiviral agent is believed to minimize the adverse effects of certain herpesvirus infections (e.g., as discussed in PCT Application WO 93 04683). Reactivating agents include agents such as estrogen, phorbol esters, forskolin and  $\beta$ -adrenergic blocking agents.

10

Agents useful for treatment of herpesvirus infections and for treatment of herpesvirus-induced KS are described in numerous U.S. Patents. For example, ganciclovir is an example of a antiviral guanine acyclic nucleotide of the type described in US Patent Nos. 4,355,032 and 4,603,219.

15

Acyclovir is an example of a class of antiviral purine derivatives, including 9-(2-hydroxyethylmethyl)adenine, of the type described in U.S. Pat. Nos. 4,267,168, 4,294,831 and 4,199,574.

20

Brivudin is an example of an antiviral deoxyuridine derivative of the type described in US Patent No. 4,424,211.

25

Vidarabine is an example of an antiviral purine nucleoside of the type described in British Pat. 1,159,290.

30

Brovavir is an example of an antiviral deoxyuridine derivative of the type described in US Patent Nos. 4,542,210 and 4,386,076.

35

BHCG is an example of an antiviral carbocyclic nucleoside analogue of the type described in US Patent Nos. 5,153,352, 5,034,394 and 5,126,345.

HPMPC is an example of an antiviral phosphonyl methoxyalkyl derivative with of the type described in US Patent No. 5,142,051.

5 CDG (Carbocyclic 2'-deoxyguanosine) is an example of an antiviral carbocyclic nucleoside analogue of the type described in US Patent Nos. 4,543,255, 4,855,466, and 4,894,458.

10 Foscarnet is described in US Patent No. 4,339,445.

Trifluridine and its corresponding ribonucleoside is described in US Patent No. 3,201,387.

15 U.S. Patent No. 5,321,030 (Kaddurah-Daouk et al.; Amira) describes the use of creatine analogs as antiherpes viral agents. U.S. Patent No. 5,306,722 (Kim et al.; Bristol-Meyers Squibb) describes thymidine kinase inhibitors useful for treating HSV  
20 infections and for inhibiting herpes thymidine kinase. Other antiherpesvirus compositions are described in U.S. Patent Nos. 5,266,649 and 5,098,708 (Konishi et al.; Bristol-Meyers Squibb) and 5,175,165 (Blumenkopf et al.; Burroughs Wellcome). U.S. Patent No.  
25 4,860,820 (Ashton et al.; Merck) describes the antiherpes virus agent (S)-9-(2,3-dihydroxy-1-propoxymethyl)guanine.

U.S. Patent No. 4,708,935 (Suhadolnik et al.; Research  
30 Corporation) describes a 3'-deoxyadenosine compound effective in inhibiting HSV and EBV. U.S. Patent No. 4,386,076 (Machida et al.; Yamasa Shoyu Kabushiki Kaisha) describes use of  
(E)-5-(2-halogenovinyl)-arabino-furanosyluracil as an  
35 antiherpesvirus agent. U.S. Patent No. 4,340,599 (Lieb et al.; Bayer Aktiengesellschaft) describes phosphonohydroxyacetic acid derivatives useful as

adjuvants; or nontoxic, nontherapeutic, nonimmunogenic stabilizers and the like. Effective amounts of such diluent or carrier are those amounts which are effective to obtain a pharmaceutically acceptable formulation in terms of solubility of components, or biological activity, etc.

#### V. Immunological Approaches to Therapy

Having identified a primary causal agent of KS in humans as a novel human herpesvirus, there are immunosuppressive therapies that can modulate the immunologic dysfunction that arises from the presence of viral-infected tissue. In particular, agents that block the immunological attack of the viral-infected cells will ameliorate the symptoms of KS and/or reduce disease progression. Such therapies include antibodies that prevent immune system targeting of viral-infected cells. Such agents include antibodies which bind to cytokines that otherwise upregulate the immune system in response to viral infection.

The antibody may be administered to a patient either singly or in a cocktail containing two or more antibodies, other therapeutic agents, compositions, or the like, including, but not limited to, immunosuppressive agents, potentiators and side-effect relieving agents. Of particular interest are immunosuppressive agents useful in suppressing allergic reactions of a host. Immunosuppressive agents of interest include prednisone, prednisolone, DECADRON (Merck, Sharp & Dohme, West Point, PA), cyclophosphamide, cyclosporine, 6-mercaptopurine, methotrexate, azathioprine and i.v. gamma globulin or their combination. Potentiators of interest include monensin, ammonium chloride and chloroquine. All of these agents are administered in generally accepted

efficacious dose ranges such as those disclosed in the Physician Desk Reference, 41st Ed. (1987), Publisher Edward R. Barnhart, New Jersey.

5 Immune globulin from persons previously infected with human herpesviruses or related viruses can be obtained using standard techniques. Appropriate titers of antibodies are known for this therapy and are readily applied to the treatment of KS. Immune globulin can  
10 be administered via parenteral injection or by intrathecal shunt. In brief, immune globulin preparations may be obtained from individual donors who are screened for antibodies to the KS-associated human herpesvirus, and plasmas from high-titered  
15 donors are pooled. Alternatively, plasmas from donors are pooled and then tested for antibodies to the human herpesvirus of the invention; high-titered pools are then selected for use in KS patients.

20 Antibodies may be formulated into an injectable preparation. Parenteral formulations are known and are suitable for use in the invention, preferably for i.m. or i.v. administration. The formulations containing therapeutically effective amounts of  
25 antibodies or immunotoxins are either sterile liquid solutions, liquid suspensions or lyophilized versions and optionally contain stabilizers or excipients. Lyophilized compositions are reconstituted with suitable diluents, e.g., water for injection, saline,  
30 0.3% glycine and the like, at a level of about from .01 mg/kg of host body weight to 10 mg/kg where appropriate. Typically, the pharmaceutical compositions containing the antibodies or immunotoxins will be administered in a therapeutically effective  
35 dose in a range of from about .01 mg/kg to about 5 mg/kg of the treated mammal. A preferred therapeutically effective dose of the pharmaceutical

composition containing antibody or immunotoxin will be in a range of from about 0.01 mg/kg to about 0.5 mg/kg body weight of the treated mammal administered over several days to two weeks by daily intravenous infusion, each given over a one hour period, in a sequential patient dose-escalation regimen.

Antibody may be administered systemically by injection i.m., subcutaneously or intraperitoneally or directly into KS lesions. The dose will be dependent upon the properties of the antibody or immunotoxin employed, e.g., its activity and biological half-life, the concentration of antibody in the formulation, the site and rate of dosage, the clinical tolerance of the patient involved, the disease afflicting the patient and the like as is well within the skill of the physician.

The antibody of the present invention may be administered in solution. The pH of the solution should be in the range of pH 5 to 9.5, preferably pH 6.5 to 7.5. The antibody or derivatives thereof should be in a solution having a suitable pharmaceutically acceptable buffer such as phosphate, tris (hydroxymethyl) aminomethane-HCl or citrate and the like. Buffer concentrations should be in the range of 1 to 100 mM. The solution of antibody may also contain a salt, such as sodium chloride or potassium chloride in a concentration of 50 to 150 mM. An effective amount of a stabilizing agent such as an albumin, a globulin, a gelatin, a protamine or a salt of protamine may also be included and may be added to a solution containing antibody or immunotoxin or to the composition from which the solution is prepared.

Systemic administration of antibody is made daily, generally by intramuscular injection, although



intravascular infusion is acceptable. Administration may also be intranasal or by other nonparenteral routes. Antibody or immunotoxin may also be administered via microspheres, liposomes or other microparticulate delivery systems placed in certain tissues including blood.

In therapeutic applications, the dosages of compounds used in accordance with the invention vary depending on the class of compound and the condition being treated. The age, weight, and clinical condition of the recipient patient; and the experience and judgment of the clinician or practitioner administering the therapy are among the factors affecting the selected dosage. For example, the dosage of an immunoglobulin can range from about 0.1 milligram per kilogram of body weight per day to about 10 mg/kg per day for polyclonal antibodies and about 5% to about 20% of that amount for monoclonal antibodies. In such a case, the immunoglobulin can be administered once daily as an intravenous infusion. Preferably, the dosage is repeated daily until either a therapeutic result is achieved or until side effects warrant discontinuation of therapy. Generally, the dose should be sufficient to treat or ameliorate symptoms or signs of KS without producing unacceptable toxicity to the patient.

An effective amount of the compound is that which provides either subjective relief of a symptom(s) or an objectively identifiable improvement as noted by the clinician or other qualified observer. The dosing range varies with the compound used, the route of administration and the potency of the particular compound.

## VI. Vaccines and Prophylaxis for KS

This invention provides substances suitable for use as vaccines for the prevention of KS and methods for administering them. The vaccines are directed against KSHV and most preferably comprise antigens obtained from KSHV. In one embodiment, the vaccine contains attenuated KSHV. In another embodiment, the vaccine contains killed KSHV. In another embodiment, the vaccine contains a nucleic acid vector encoding a KSHV polypeptide. In another embodiment, the vaccine is a subunit vaccine containing a KSHV polypeptide.

This invention provides a recombinant KSHV virus with a gene encoding a KSHV polypeptide deleted from the genome. The recombinant virus is useful as an attenuated vaccine to prevent KSHV infection.

This invention provides a method of vaccinating a subject against Kaposi's sarcoma, comprising administering to the subject an effective amount of the peptide or polypeptide encoded by the isolated DNA molecule, and a suitable acceptable carrier, thereby vaccinating the subject. In one embodiment naked DNA is administered to the subject in an effective amount to vaccinate the subject against Kaposi's sarcoma.

This invention provides a method of immunizing a subject against disease caused by KSHV which comprises administering to the subject an effective immunizing dose of an isolated herpesvirus subunit vaccine.

### A. Vaccines

The vaccine can be made using synthetic peptide or recombinantly-produced polypeptide described above as antigen. Typically, a vaccine will include from about

1 to 50 micrograms of antigen. More preferably, the amount of polypeptide is from about 15 to about 45 micrograms. Typically, the vaccine is formulated so that a dose includes about 0.5 milliliters. The vaccine may be administered by any route known in the art. Preferably, the route is parenteral. More preferably, it is subcutaneous or intramuscular.

There are a number of strategies for amplifying an antigen's effectiveness, particularly as related to the art of vaccines. For example, cyclization or circularization of a peptide can increase the peptide's antigenic and immunogenic potency. See U.S. Pat. No. 5,001,049. More conventionally, an antigen can be conjugated to a suitable carrier, usually a protein molecule. This procedure has several facets. It can allow multiple copies of an antigen, such as a peptide, to be conjugated to a single larger carrier molecule. Additionally, the carrier may possess properties which facilitate transport, binding, absorption or transfer of the antigen.

For parenteral administration, such as subcutaneous injection, examples of suitable carriers are the tetanus toxoid, the diphtheria toxoid, serum albumin and lamprey, or keyhole limpet, hemocyanin because they provide the resultant conjugate with minimum genetic restriction. Conjugates including these universal carriers can function as T cell clone activators in individuals having very different gene sets.

The conjugation between a peptide and a carrier can be accomplished using one of the methods known in the art. Specifically, the conjugation can use bifunctional cross-linkers as binding agents as detailed, for example, by Means and Feeney, "A recent

polypeptides from the human herpesvirus. For example, it is known in the protein art that certain amino acid residues can be substituted with amino acids of similar size and polarity without an undue effect upon the biological activity of the protein. The human herpesvirus polypeptides have significant tertiary structure and the epitopes are usually conformational. Thus, modifications should generally preserve conformation to produce a protective immune response.

10

#### B. Antibody Prophylaxis

Therapeutic, intravenous, polyclonal or monoclonal antibodies can be used as a mode of passive immunotherapy of herpesviral diseases including perinatal varicella and CMV. Immune globulin from persons previously infected with the human herpesvirus and bearing a suitably high titer of antibodies against the virus can be given in combination with antiviral agents (e.g. ganciclovir), or in combination with other modes of immunotherapy that are currently being evaluated for the treatment of KS, which are targeted to modulating the immune response (i.e. treatment with copolymer-1, antiidiotypic monoclonal antibodies, T cell "vaccination"). Antibodies to human herpesvirus can be administered to the patient as described herein. Antibodies specific for an epitope expressed on cells infected with the human herpesvirus are preferred and can be obtained as described above.

30

A polypeptide, analog or active fragment can be formulated into the therapeutic composition as neutralized pharmaceutically acceptable salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide or antibody molecule) and which are

35

The vaccines may be administered by any conventional method for the administration of vaccines including oral and parenteral (e.g., subcutaneous or intramuscular) injection. Intramuscular administration is preferred. The treatment may consist of a single dose of vaccine or a plurality of doses over a period of time. It is preferred that the dose be given to a human patient within the first 8 months of life. The antigen of the invention can be combined with appropriate doses of compounds including influenza antigens, such as influenza type A antigens. Also, the antigen could be a component of a recombinant vaccine which could be adaptable for oral administration.

Vaccines of the invention may be combined with other vaccines for other diseases to produce multivalent vaccines. A pharmaceutically effective amount of the antigen can be employed with a pharmaceutically acceptable carrier such as a protein or diluent useful for the vaccination of mammals, particularly humans. Other vaccines may be prepared according to methods well-known to those skilled in the art.

Those of skill will readily recognize that it is only necessary to expose a mammal to appropriate epitopes in order to elicit effective immunoprotection. The epitopes are typically segments of amino acids which are a small portion of the whole protein. Using recombinant genetics, it is routine to alter a natural protein's primary structure to create derivatives embracing epitopes that are identical to or substantially the same as (immunologically equivalent to) the naturally occurring epitopes. Such derivatives may include peptide fragments, amino acid substitutions, amino acid deletions and amino acid additions of the amino acid sequence for the viral

formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed from the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

### C. Monitoring Therapeutic Efficacy

This invention provides a method for monitoring the therapeutic efficacy of treatment for Kaposi's sarcoma which comprises: (a) determining in a first sample from a subject with Kaposi's sarcoma the presence of the isolated nucleic acid molecule; (b) administering to the subject a therapeutic amount of an agent such that the agent is contacted to the cell in a sample; (c) determining after a suitable period of time the amount of the isolated nucleic acid molecule in the second sample from the treated subject; and (d) comparing the amount of isolated nucleic acid molecule determined in the first sample with the amount determined in the second sample, a difference indicating the effectiveness of the agent, thereby monitoring the therapeutic efficacy of treatment for Kaposi's sarcoma. As defined herein "amount" is viral load or copy number. Methods of determining viral load or copy number are known to those skilled in the art.

### VII. Screening Assays For Pharmaceuticals for Alleviating the Symptoms of KS

Since an agent involved in the causation or progression of KS has been identified and described,

assays directed to identifying potential pharmaceutical agents that inhibit the biological activity of the agent are possible. KS drug screening assays which determine whether or not a drug has activity against the virus described herein are contemplated in this invention. Such assays comprise incubating a compound to be evaluated for use in KS treatment with cells which express the KS associated human herpesvirus polypeptides or peptides and determining therefrom the effect of the compound on the activity of such agent. In vitro assays in which the virus is maintained in suitable cell culture are preferred, though in vivo animal models would also be effective.

Compounds with activity against the agent of interest or peptides from such agent can be screened in in vitro as well as in vivo assay systems. In vitro assays include infecting peripheral blood leukocytes or susceptible T cell lines such as MT-4 with the agent of interest in the presence of varying concentrations of compounds targeted against viral replication, including nucleoside analogs, chain terminators, antisense oligonucleotides and random polypeptides (Asada et al., 1989, *J. Clin. Microbiol.* 27, 2204; Kikuta et al., 1989, *Lancet* Oct. 7, 861). Infected cultures and their supernatants can be assayed for the total amount of virus including the presence of the viral genome by quantitative PCR, by dot blot assays or by using immunologic methods. For example, a culture of susceptible cells could be infected with KSHV in the presence of various concentrations of drug, fixed on slides after a period of days, and examined for viral antigen by indirect immunofluorescence with monoclonal antibodies to viral polypeptides (Kikuta et al., *supra*). Alternatively, chemically adhered MT-4 cell monolayers can be used

review of protein modification techniques.  
*Bioconjugate Chem.* 1, 2-12 (1990).

Vaccines against a number of the Herpesviruses have  
5 been successfully developed. Vaccines against  
Varicella-Zoster Virus using a live attenuated Oka  
strain is effective in preventing herpes zoster in the  
elderly, and in preventing chickenpox in both  
immunocompromised and normal children (Hardy, I., et  
10 al., 1990, *Inf. Dis. Clin. N. Amer.* 4, 159; Hardy, I.  
et al., 1991, *New Engl. J. Med.* 325, 1545; Levin, M.J.  
et al., 1992, *J. Inf. Dis.* 166, 253; Gershon, A.A.,  
1992, *J. Inf. Dis.* 166(Suppl), 563. Vaccines against  
Herpes simplex Types 1 and 2 are also commercially  
15 available with some success in protection against  
primary disease, but have been less successful in  
preventing the establishment of latent infection in  
sensory ganglia (Roizman, B., 1991, *Rev. Inf. Disease*  
13(Suppl. 11), S892; Skinner, G.R. et al., 1990, *Med.*  
20 *Microbiol. Immunol.* 180, 305).

Vaccines against KSHV can be made from the KSHV  
envelope glycoproteins. These polypeptides can be  
purified and used for vaccination (Lasky, L.A., 1990,  
25 *J. Med. Virol.* 31, 59). MHC-binding peptides from  
cells infected with the human herpesvirus can be  
identified for vaccine candidates per the methodology  
of Marloes, et al., 1991, *Eur. J. Immunol.* 21, 2963-  
2970.

30 The KSHV antigen may be combined or mixed with various  
solutions and other compounds as is known in the art.  
For example, it may be administered in water, saline  
or buffered vehicles with or without various adjuvants  
or immunodiluting agents. Examples of such adjuvants  
35 or agents include aluminum hydroxide, aluminum  
phosphate, aluminum potassium sulfate (alum),



beryllium sulfate, silica, kaolin, carbon, water-in-oil emulsions, oil-in-water emulsions, muramyl dipeptide, bacterial endotoxin, lipid X, *Corynebacterium parvum* (*Propionibacterium acnes*, *Bordetella pertussis*, polyribonucleotides, sodium alginate, lanolin, lysolecithin, vitamin A, saponin, liposomes, levamisole, DEAE-dextran, blocked copolymers or other synthetic adjuvants. Such adjuvants are available commercially from various sources, for example, Merck Adjuvant 65 (Merck and Company, Inc., Rahway, N.J.) or Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, Michigan). Other suitable adjuvants are Amphigen (oil-in-water), Alhydrogel (aluminum hydroxide), or a mixture of Amphigen and Alhydrogel. Only aluminum is approved for human use.

The proportion of antigen and adjuvant can be varied over a broad range so long as both are present in effective amounts. For example, aluminum hydroxide can be present in an amount of about 0.5% of the vaccine mixture ( $Al_2O_3$  basis). On a per-dose basis, the amount of the antigen can range from about 0.1  $\mu g$  to about 100  $\mu g$  protein per patient. A preferable range is from about 1  $\mu g$  to about 50  $\mu g$  per dose. A more preferred range is about 15  $\mu g$  to about 45  $\mu g$ . A suitable dose size is about 0.5 ml. Accordingly, a dose for intramuscular injection, for example, would comprise 0.5 ml containing 45  $\mu g$  of antigen in admixture with 0.5% aluminum hydroxide. After formulation, the vaccine may be incorporated into a sterile container which is then sealed and stored at a low temperature, for example 4°C, or it may be freeze-dried. Lyophilization permits long-term storage in a stabilized form.

for an infectious agent assay using indirect immunofluorescent antibody staining to search for focus reduction (Higashi et al., 1989, J. Clin. Microb. 27, 2204).

5 As an alternative to whole cell in vitro assays, purified KSHV enzymes isolated from a host cell or produced by recombinant techniques can be used as targets for rational drug design to determine the effect of the potential drug on enzyme activity. KSHV  
10 enzymes amenable to this approach include, but are not limited to, dihydrofolate reductase (DHFR), thymidylate synthase (TS), thymidine kinase or DNA polymerase. A measure of enzyme activity indicates  
15 effect on the agent itself.

Drug screens using herpes viral products are known and have been previously described in EP 0514830 (herpes proteases) and WO 94/04920 (UL13 gene product).

20 This invention provides an assay for screening anti-KS chemotherapeutics. Infected cells can be incubated in the presence of a chemical agent that is a potential chemotherapeutic against KS (e.g., acyclo-guanosine).  
25 The level of virus in the cells is then determined after several days by immunofluorescence assay for antigens, Southern blotting for viral genome DNA or Northern blotting for mRNA and compared to control cells. This assay can quickly screen large numbers of  
30 chemical compounds that may be useful against KS.

Further, this invention provides an assay system that is employed to identify drugs or other molecules capable of binding to the nucleic acid molecule or  
35 proteins, either in the cytoplasm or in the nucleus, thereby inhibiting or potentiating transcriptional activity. Such assay would be useful in the

development of drugs that would be specific against particular cellular activity, or that would potentiate such activity, in time or in level of activity.

5 This invention provides a method of screening for a KSHV-selective antiviral drug in vivo comprising: (a) expression of KSHV DHFR or KSHV TS in a bacterial auxotroph (nutritional mutant); (b) measuring bacterial growth rate in the absence and presence of  
10 the drug; and (c) comparing the rates so measured so as to identify the drug that inhibits KSHV DHFR or KSHV TS in vivo.

15 Methods well known to those skilled in the art allow selection or production of a suitable bacterial auxotroph and measurement of bacterial growth.

The following reviews of antifolate compounds are provided to more fully describe the state of the art, particularly as it pertains to inhibitors of  
20 dihydrofolate reductase and thymidylate synthase: (a) Unger, 1996, Current concepts of treatment in medical oncology: new anticancer drugs, *Journal of Cancer Research & Clinical Oncology* 122, 189-198; (b)  
25 Jackson, 1995, Toxicity prediction from metabolic pathway modelling, *Toxicology* 102, 197-205; (c) Schultz, 1995, Newer antifolates in cancer therapy, *Progress in Drug Research* 44, 129-157; (d) van der Wilt and Peters, 1994, New targets for pyrimidine  
30 antimetabolites in the treatment of solid tumours 1: Thymidylate synthase, *Pharm World Sci* 16, 167; (e) Fleisher, 1993, Antifolate analogs: mechanism of action, analytical methodology, and clinical efficacy, *Therapeutic Drug Monitoring* 15, 521-526; (f) Eggott et  
35 al., 1993, Antifolates in rheumatoid arthritis: a hypothetical mechanism of action, *Clinical & Experimental Rheumatology* 11 Suppl 8, S101-S105; (g)

Huennekens et al., 1992, Membrane transport of folate compounds, *Journal of Nutritional Science & Vitaminology* Spec No, 52-57; (ii) Fleming and Schilsky, 1992, Antifolates: the next generation, *Seminars in Oncology* 19, 707-719; and (i) Bertino et al., 1992, Enzymes of the thymidylate cycle as targets for chemotherapeutic agents: mechanisms of resistance, *Mount Sinai Journal of Medicine* 59, 391-395.

10 This invention provides a method of determining the health of a subject with AIDS comprising: (a) measuring the plasma concentration of vMIP-I, vMIP-II or vMIP-III; and (b) comparing the measured value to a standard curve relating AIDS clinical course to the measured value so as to determine the health of the subject.

#### VIII. Treatment of HIV

20 This invention provides a method of inhibiting HIV replication, comprising administering to the subject or treating cells of a subject with an effective amount of a polypeptide which is encoded by a nucleic acid molecule, so as to inhibit replication of HIV. In one embodiment, the polypeptide is one from the list provided in Table I.

30 This invention is further illustrated in the Experimental Details Sections which follow. These sections are set forth to aid in understanding the invention but is not intended to, and should not be construed to, limit in any way the invention as set forth in the claims which follow thereafter.

#### EXPERIMENTAL DETAILS SECTION I

NUCLEOTIDE SEQUENCE OF THE KAPOSI'S SARCOMA-ASSOCIATED  
HERPESVIRUS

The genome of the Kaposi's sarcoma-associated  
5 herpesvirus (KSHV or HHV8) was mapped with cosmid and  
phage genomic libraries from the BC-1 cell line. Its  
nucleotide sequence was determined except for a 3 kb  
region at the right end of the genome that was  
refractory to cloning. The BC-1 KSHV genome consists  
10 of a 140.5 kb long unique coding region (LUR) flanked  
by multiple G-C rich 801 bp terminal repeat sequences.  
A genomic duplication that apparently arose in the  
parental tumor is present in this cell culture-derived  
strain. At least 81 open reading frames (ORFs),  
15 including 66 with similarity to herpesvirus saimiri  
ORFs, and 5 internal repeat regions are present in the  
LUR. The virus encodes genes similar to  
complement-binding proteins, three cytokines (two  
macrophage inflammatory proteins and interleukin-6),  
20 dihydrofolate reductase, bcl-2, interferon regulatory  
factor, IL-8 receptor, NCAM-like adhesin, and a D-type  
cyclin, as well as viral structural and metabolic  
proteins. Terminal repeat analysis of virus DNA from  
a KS lesion suggests a monoclonal expansion of KSHV in  
25 the KS tumor. The complete genome sequence is set  
forth in Genbank Accession Numbers U75698 (LUR),  
U75699 (TR) and U75700 (ITR).

Kaposi's sarcoma is a vascular tumor of mixed cellular  
30 composition (Tappero et al., 1993, *J. Am. Acad.  
Dermatol.* 28, 371-395). The histology and relatively  
benign course in persons without severe  
immunosuppression has led to suggestions that KS tumor  
cell proliferation is cytokine induced (Ensoli et al.,  
35 1992, *Immunol. Rev.* 127, 147-155). Epidemiologic  
studies indicate the tumor is under strict immunologic  
control and is likely to be caused by a sexually

transmitted infectious agent other than HIV (Peterman et al., 1993, *AIDS* 7, 605-611). KS-associated herpesvirus (KSHV) was discovered in an AIDS-KS lesion by representational difference analysis (RDA) and shown to be present in almost all AIDS-KS lesions (Chang et al., 1994, *Science* 265, 1865-1869). These findings have been confirmed and extended to nearly all KS lesions examined from the various epidemiologic classes of KS (Boshoff et al., 1995, *Lancet* 345, 1043-1044; Dupin et al., 1995, *Lancet* 345, 761-762; Moore and Chang, 1995, *New Eng. J. Med.* 332, 1181-1185; Schalling et al., 1995, *Nature Med.* 1, 707-708; Chang et al., 1996, *Arch. Int. Med.* 156, 202-204). KSHV is the eighth presumed human herpesvirus (HHV8) identified to date.

The virus was initially identified from two herpesvirus DNA fragments, KS330Bam and KS631Bam (Chang et al., 1994, *Science* 265, 1865-1869). Subsequent sequencing of a 21 kb AIDS-KS genomic library fragment (KS5) hybridizing to KS330Bam demonstrated that KSHV is a gammaherpesvirus related to herpesvirus saimiri (HVS) belonging to the genus Rhadinovirus (Moore et al., 1996, *J. Virol.* 70, 549-558). Colinear similarity (synteny) of genes in this region is maintained between KSHV and HVS, as well as Epstein-Barr virus (EBV) and equine herpesvirus 2 (EHV2). A 12 kb region (L54 and SGL-1) containing the KS631Bam sequence includes cyclin D and IL-8Ra genes unique to rhadinoviruses.

KSHV is not readily transmitted to uninfected cell lines (Moore et al., 1996, *J. Virol.* 70, 549-558), but it is present in a rare B cell primary effusion (body cavity-based) lymphoma (PEL) frequently associated with KS (Cesarman et al., 1995, *New Eng. J. Med.* 332, 1186-1191). BC-1 is a PEL cell line

containing a high KSHV genome copy number and is  
coinfecting with EBV (Cesarman et al., 1995, *Blood* 86,  
2708-2714). The KSHV genome form in BC-1 and its  
parental tumor comigrates with 270 kb linear markers  
on pulsed field gel electrophoresis (PFGE). (Moore et  
5 al., 1996, *J. Virol.* 70, 549-558). However, the  
genome size based on encapsidated DNA from an  
EBV-negative cell line (Renne et al., 1996, *Nature*  
*Med.* 2, 342-346) is estimated to be 165 kb (Moore et  
10 al., 1996, *J. Virol.* 70, 549-558). Estimates from KS  
lesions indicate a genome size larger than that of EBV  
(172 kb) (Decker et al., 1996, *J. Exp. Med.* 184,  
283-288).

15 To determine the genomic sequence of KSHV and identify  
novel virus genes, contiguous overlapping virus DNA  
inserts from BC-1 genomic libraries were mapped. With  
the exception of a small, unclonable repeat region at  
its right end, the genome was sequenced to high  
20 redundancy allowing definition of the viral genome  
structure and identification of genes that may play a  
role in KSHV-related pathogenesis.

#### MATERIALS AND METHODS

25 Library generation and screening. BC-1, HBL-6 and  
BCP-1 cells were maintained in RPMI 1640 with 20%  
fetal calf serum (Moore et al., 1996, *J. Virol.* 70,  
549-558; Cesarman et al., 1995, *Blood* 86, 2708-2714;  
30 Gao et al., 1996, *Nature Med.* 2, 925-928). DNA from  
BC-1 cells was commercially cloned (Sambrook et al.,  
1989, *Molecular Cloning: A laboratory manual*, Cold  
Spring Harbor Press, Salem, Mass.) into either Lambda  
FIX II or S-Cos1 vectors (Stratagene, La Jolla, CA).  
35 Phage and cosmid libraries were screened by standard  
methods (Benton et al., 1977, *Science* 196, 180-182;

Hanahan and Meselson, 1983, *Methods Enzymol.* 105, 333-342).

5 Initial library screening was performed using the  
KS330Bam and KS631Bam RDA fragments (Chang et al.,  
1994, *Science* 265, 1863-1869). Overlapping clones  
were sequentially identified using probes synthesized  
from the ends of previously identified clones (Figure  
1) (Feinberg and Vogelstein, 1983, *Anal. Biochem.* 132,  
10 6; Melton et al., 1984, *Nucl. Acids Res.* 12,  
7035-7056). The map was considered circularly  
permuted by the presence of multiple, identical TR  
units in cosmids Z2 and Z6. Each candidate phage or  
cosmid was confirmed by tertiary screening.

15

#### Shotgun sequencing and sequence verification

Lambda and cosmid DNA was purified by standard methods  
(Sambrook et al., 1989, *Molecular Cloning: A*  
20 *laboratory manual*, Cold Spring Harbor Press, Salem,  
Mass.). Shotgun sequencing (Deininger, 1983, *Anal.*  
*Biochem.* 129, 216-223; Bankier et al., 1987, *Meth.*  
*Enzymol.* 155, 51-93) was performed on sonicated DNA.  
A 1-4 kb fraction was subcloned into M13mp19 (New  
25 England Biolabs, Inc., Beverly, MA) and propagated in  
XL1-Blue cells (Stratagene, La Jolla, CA) (Sambrook et  
al., 1989, *Molecular Cloning: A laboratory manual*,  
Cold Spring Harbor Press, Salem, Mass.) M13 phages  
were positively screened using insert DNA from the  
30 phage or cosmid, and negatively screened with vector  
arm DNA or adjacent genome inserts.

Automated dideoxy cycle sequencing was performed with  
M13 (-21) CS+ or FS dye primer kits (Perkin-Elmer,  
35 Branchburg NJ) on ABI 373A or 377 sequencers (ABI,  
Foster City, CA). Approximately 300 M13 sequences  
were typically required to achieve initial coverage



for each 10 kb of insert sequence. Minimum sequence fidelity standards were defined as complete bidirectional coverage with at least 4 overlapping sequences at any given site. For regions with  
5 sequence gaps, ambiguities or frameshifts that did not meet these criteria, primer walking was done with custom primers (Perkin-Elmer) and dye terminator chemistry (FS or Ready Reaction kits, Perkin-Elmer). An unsequenced 3 kb region adjacent to the right end  
10 TR sequence in the Z2 cosmid insert could not be cloned into M13 or Bluescript despite repeated efforts.

#### Sequence assembly and open reading frame analysis

15 Sequence data were edited using Factura (ABI, Foster City, CA) and assembled into contiguous sequences using electropherograms with AutoAssembler (ABI, Foster City, CA) and into larger assemblies with  
20 AssemblyLIGN (IBI-Kodak, Rochester NY). Base positions not clearly resolved by multiple sequencing attempts (less than 10 bases in total) were assigned the majority base pair designation. The entire  
25 sequence (in 1-5 kb fragments) and all predicted open reading frames (ORFs) were analyzed using BLASTX, BLASTP and BLASTN (Altschul et al., 1990, J. Mol. Biol. 215, 403-410). The sequence was further  
30 analyzed using MOTIFS (Moore et al., 1996, J. Virol. 70, 549-558), REPEAT and BESTFIT (GCG), and MacVector (IBI, New Haven, CT).

#### ORF assignment and nomenclature

35 All ORFs with similarities to HVS were identified. These and other potential ORFs having >100 amino acids were found using MacVector. ORFs not similar to HVS ORFs were included in the map (Fig. 1) based on

similarity to other known genes, optimum initiation codon context (Kozak, 1987, Nucl. Acids Res. 15, 8125-8148), size and position. Conservative selections were made to minimize spurious assignments; this underestimates the number of true reading frames. KSHV ORF nomenclature is based on HVS similarities; KSHV ORFs not similar to HVS genes are numbered in consecutive order with a K prefix. ORFs with sequence but not positional similarity to HVS ORFs were assigned the HVS ORF number (e.g., ORF 2). As new ORFs are identified, it is suggested that they be designated by decimal notation. The standard map orientation (Fig. 1) of the KSHV genome is the same as for HVS (Albrecht et al., 1992, J. Virol. 66, 5047-5058) and EBV2 (Telford et al., 1995, J. Mol. Biol. 249, 520-528), and reversed relative to the EBV standard map (Baer et al., 1984, Nature 310, 207-211).

## RESULTS

20

### Genomic mapping and sequence characteristics

Complete genome mapping was achieved with 7 lambda and 3 cosmid clones (Fig. 1). The structure of the BC-1 KSHV genome is similar to HVS in having a long unique region (LUR) flanked by TR units. The ~140.5 kb LUR sequence has 53.5% G+C content and includes all identified KSHV ORFs. TR regions consist of multiple 801 bp direct repeat units having 84.5% G-C content (Fig. 2A) with potential packaging and cleavage sites. Minor sequence variations are present among repeat units. The first TR unit at the left (Z6) TR junction (205bp) is deleted and truncated in BC-1 compared to the prototypical TR unit.

35

The genome sequence abutting the right terminal repeat region is incomplete due to a 3 kb region in the Z2

cosmid insert that could not be cloned into sequencing vectors. Partial sequence information from primer walking indicates that this region contains stretches of 16 bp A-G rich imperfect direct repeats interspersed with at least one stretch of 16 bp C-T rich imperfect direct repeats. These may form a larger inverted repeat that could have contributed to our difficulty in subcloning this region. Greater than 12-fold average sequence redundancy was achieved for the entire LUR with complete bidirectional coverage by at least 4 overlapping reads except in the unclonable region.

The BC-1 TR region was examined by Southern blotting since sequencing of the entire region is not possible due to its repeat structure. BC-1, BCP-1 (an EBV-negative, KSHV infected cell line) and KS lesion DNAs have an intense ~800 bp signal consistent with the unit length repeat sequence when digested with enzymes that cut once in the TR and hybridized to a TR probe (Figs. 2B and 2C). Digestion with enzymes that do not cut in the TR indicates that the BC-1 strain contains a unique region buried in the TR, flanked by ~7 kb and ~35 kb TR sequences (Figs. 2C and 2D). An identical pattern occurs in HBL-6, a cell line independently derived from the same tumor as BC-1, suggesting that this duplication was present in the parental tumor (Figs. 2C and 2D). The restriction pattern with Not I, which also cuts only once within the TR but rarely within the LUR, suggests that the buried region is at least 33 kb. Partial sequencing of this region demonstrates that it is a precise genomic duplication of the region beginning at ORF X8. The LUR is 140 kb including the right end unsequenced gap (<3kb). The estimated KSHV genomic size in BC-1 and HBL-6 (including the duplicated region) is approximately 210 kb.

Based on the EBV replication model used in clonality studies (Raab-Traub and Flynn, 1986, Cell 47, 883-889), the polymorphic BCP-1 ladder pattern may reflect lytic virus replication and superinfection (Fig. 2C). The EBV ladder pattern occurs when TR units are deleted or duplicated during lytic replication and is a stochastic process for each infected cell (Raab-Traub and Flynn, 1986, Cell 47, 883-889). No ladder pattern is present for BC-1 which is under tight latent KSHV replication control (Moore et al., 1996, J. Virol. 70, 549-558). KS lesion DNA also shows a single hybridizing band suggesting that virus in KS tumor cells may be of monoclonal origin.

#### 15 Features and coding regions of the KSHV LUR

The KSHV genome shares the 7 block (B) organization (B1-B7, Fig. 1) of other herpesviruses (Chee et al., 1990, Curr. Topics Microbiol. Immunol. 154, 135-169), with sub-family specific or unique ORFs present between blocks (interblock regions (IB) a-h, Fig. 1). ORF analysis indicates that only 79% of the sequenced 137.5 kb LUR encodes 81 identifiable ORFs which is likely to be due to a conservative assignment of ORF positions. The overall LUR CpG dinucleotide observed/expected (O/E) ratio is 0.75 consistent with a moderate loss of methylated cytosines, but there is marked regional variation. The lowest CpG O/E ratios (<0.67) occur in IBa (bp 1-3200), in B5 (68,602-69,405) and IBh (117,352-137,507). The highest O/E ratios (>0.88) extend from B2 to B3 (30,701-47,849), in IBc (67,301-68,600), and in B6 (77,251-83,600). Comparison to the KSS sequence (Moore et al., 1996, J. Virol. 70, 549-558) shows a high sequence conservation between these two strains with only 21 point mutations over the comparable 20.7 kb region (0.1%). A frameshift within BC-1 ORF 28

(position 49,004) compared to KSS ORF 28 was not resolvable despite repeated sequencing of KSS and PCR products amplified from BC-1. Two additional frameshifts in noncoding regions (bp 47,861 and 49,338) are also present compared to the KSS sequence.

Several repeat regions are present in the LUR (Fig. 1). A 143 bp sequence is repeated within ORF K11 at positions 92,678-92,820 and 92,852-92,994 (waka [wka]). Complex repeats are present in other regions of the genome: 20 and 30 bp repeats in the region from 24,285-24,902 (frnk), a 13 bp repeat between bases 29,775 and 29,942 (vnct), two separate 23 bp repeat stretches between bases 118,123 and 118,697 (pppa), and 15 different 11-16 bp repeats throughout the region from 124,527 to 126,276 (moi). A complex A-G rich repeat region (mdsk) begins at 137,099 and extends into the unsequenced gap.

Conserved ORFs with similar genes found in other herpesviruses are listed in Table 1, along with their polarity, map positions, sizes, relatedness to HVS and EBV ORFs, and putative functions. Conserved ORFs coding for viral structural proteins and enzymes include genes involved in viral DNA replication (e.g., DNA polymerase (ORF 9)), nucleotide synthesis (e.g., dihydrofolate reductase (DHFR, ORF 2)), thymidylate synthase (TS, ORF 70)), regulators of gene expression (R transactivator (LCTP, ORF50)) and 5 conserved herpesvirus structural capsid and 5 glycoprotein genes.

Several genes that are similar to HVS ORFs also have unique features. ORF 45 has sequence similarity to nuclear and transcription factors (chick nucleolin and yeast SIR3) and has an extended acidic domain typical for transactivator proteins between amino acids 90 and

115. ORF73 also has an extended acidic domain separated into two regions by a glutamine-rich sequence encoded by the mol repeat. The first region consists almost exclusively of aspartic and glutamic acid residue repeats while the second glutamic acid rich region has a repeated leucine heptad motif suggestive of a leucine zipper structure. ORF 75, a putative tegument protein, has a high level of similarity to the purine biosynthetic enzyme of *E. coli* and *D. melanogaster* N-formylglycinamide ribotide amidotransferase (FGARAT).

ORFs K3 and K5 are not similar to HVS genes but are similar to the major immediate early bovine herpesvirus type 4 (BHV4) gene IE1 (12 and 13% identity respectively) (van Santen, 1991, *J. Virol.* 65, 5211-5224). These genes have no significant similarity to the herpes simplex virus 1 (HSV1)  $\alpha$ 0 (which is similar to BHV4 IE1), but encode proteins sharing with the HSV1 ICP0 protein a cysteine-rich region which may form a zinc finger motif (van Santen, 1991, *J. Virol.* 65, 5211-5224). The protein encoded by ORF K5 has a region similar to the nuclear localization site present in the late form of the BHV4 protein. ORF K6 has a purine binding motif (GLLVTGKS) in the C-terminus of the protein which is similar to a motif present in the KSHV TK (ORF21) (Moore et al., 1996, *J. Virol.* 70, 549-556).

No KSHV genes with similarity to HVS ORFs 1, 3, 5, 12, 13, 14, 15, 51 and 71 were identified in the KSHV LUR sequence. HVS ORF 1 codes for a transforming protein, responsible for HVS-induced in vitro lymphocyte transformation (Akari et al., 1996, *Virology* 218, 382-388) and has poor sequence conservation among HVS strains (Jung and Desrosiers, 1991, *J. Virol.* 65, 6953-6960; Jung and Desrosiers, 1995, *Molec. Cellular*

Biol. 15, 6506-6512). Functional KSHV genes similar to this gene may be present but were not identifiable by sequence comparison. Likewise, no KSHV genes similar to EBV latency and transformation-associated proteins (EBNA-1, EBNA-2, EBNA-LP, LMP-1, LMP-2 or gp350/220) were found despite some similarity to repeat sequences present in these genes. KSHV also does not have a gene similar to the BZLF1 EBV transactivator gene.

Several sequences were not given ORF assignments although they have characteristics of expressed genes. The sequence between bp 90,173 and 90,643 is similar to the precursor of secreted glycoprotein X (gX), encoded by a number of alphaherpesviruses (pseudorabies, EHV1), and which does not form part of the virion structure. Like the cognate gene in EHV1, the KSHV form lacks the highly-acidic carboxy terminus of the pseudorabies gene.

Two polyadenylated transcripts expressed at high copy number in BCBL-1 are present at positions 28,661-29,741 (T1.1) in IBb and 116,130-117,436 (T0.7) in IBh. T0.7 encodes a 60 residue polypeptide (ORF K12, also called Kaposin) and T1.1 (also referred to as nut-1) has been speculated to be a U RNA-like transcript.

#### Cell cycle regulation and cell signaling proteins

A number of ORFs which are either unique to KSHV or shared only with other gammaherpesviruses encode genes similar to oncoproteins and cell signaling proteins. ORF 16, similar to EBV BHRF1 and HVS ORF16, encodes a functional Bcl-2-like protein which can inhibit Bax-mediated apoptosis. ORF 72 encodes a functional cyclin D gene, also found in HVS (Nicholas et al.,

1992, Nature 355, 362-365), that can substitute for human cyclin D in phosphorylating the retinoblastoma tumor suppressor protein.

5 KSHV encodes a functionally-active IL-6 (ORF K2) and two macrophage inflammatory proteins (MIPs) (ORFs K4 and K6) which are not found in other human herpesviruses. The vIL-6 has 62% amino acid similarity to the human IL-6 and can substitute for  
10 human IL-6 in preventing mouse myeloma cell apoptosis. Both MIP-like proteins have conserved C-C dimer signatures characteristic of  $\beta$ -chemokines and near sequence identity to human MIP-1 $\alpha$  in their N-terminus regions. vMIP-1 (ORF K6) can inhibit CCR-5 dependent  
15 HIV-1 replication. An open reading frame spanning nucleotide numbers (bp) 22,529-22,195 (vMIP-III) has low conservation with MIP 1 $\beta$  (BLASTX poisson p=0.0015) but retains the C-C dimer motif. ORF K9 (vIRF1) encodes a 449 residue protein with similarity to the  
20 family of interferon regulatory factors (IRF) (David, 1995, Pharmac. Ther. 65, 149-161). It has 13.4% amino acid identity to human interferon consensus sequence binding protein and partial conservation of the IRF DNA-binding domain. Three additional open reading  
25 frames at bp 88,910-88,410 (vIRF2), bp 90,541-89,600 (vIRF3) and bp 94,127-93,636 (vIRF4) also have low similarity to IRF-like proteins (p > 0.35). No conserved interferon consensus sequences were found in this region of the genome.

30 Other genes encoding signal transduction polypeptides, which are also found in other herpesviruses, include a complement-binding protein (v-CBP, ORF 4), a neural cell adhesion molecule (NCAM)-like protein (v-adh, ORF K14) and an IL8 receptor (ORF 74). Genes similar to  
35 ORFs 4 and 74 are present in other rhadinoviruses and ORF 4 is similar to variola B19L and D12L proteins.



ORF K14 (v-adh) is similar to the rat and human OX-2 membrane antigens, various NCAMs and the poliovirus receptor-related protein PRR1. OX-2 is in turn similar to ORF U85 of human herpesviruses 6 and 7 but there is no significant similarity between the KSHV and betaherpesvirus OX-2/NCAM ORFs. Like other immunoglobulin family adhesion proteins, v-adh has V-like, C-like, transmembrane and cytoplasmic domains, and an RGD binding site for fibronectin at residues 268-270. The vIL-8R has a seven transmembrane spanning domain structure characteristic of G-protein coupled chemoattractant receptors which includes the EBV-induced EB11 protein (Birkenbach et al., 1993, J. Virol. 67, 2209-2220).

#### DISCUSSION

The full-length sequence of the KSHV genome in BC-1 cells provides the opportunity to investigate molecular mechanisms of KSHV-associated pathogenesis. The KSHV genome has standard features of rhadinovirus genomes including a single unique coding region flanked by high G+C terminal repeat regions which are the presumed sites for genome circularization. In addition to having 66 conserved herpesvirus genes involved in herpesvirus replication and structure, KSHV is unique in encoding a number of proteins mimicking cell cycle regulatory and signaling proteins.

Our estimated size of the BC-1 derived genome (210 kb including the duplicated portion) is consistent with that found using encapsidated virion DNA (Zhong et al., 1996, Proc. Natl. Acad. Sci. USA 93, 6641-6646). Genomic rearrangements are common in cultured herpesviruses (Baer et al., 1984, Nature 310, 207-211; Cha et al., 1996, J. Virol. 70, 78-83). However, the

genomic duplication present in the BC-1 KSHV probably did not arise during tissue culture passage. TR hybridization studies indicate that this insertion of a duplicated LUR fragment into the BC-1 TR is also present in KSHV from the independently derived HBL-6 cell line (Gaidano et al., 1996, Leukemia 10, 1237-40).

Despite this genomic rearrangement, the KSHV genome is well conserved within coding regions. There is less than 0.1% base pair variation between the BC-1 and the 21 kb KSS fragment isolated from a KS lesion. Higher levels of variation may be present in strains from other geographic regions or other disease conditions. Within the LUR, synteny to HVS is lost at ORFs 1 and 70 but there is concordance in all other regions conserved with HVS. Several conserved genes, such as thymidine kinase (TK) (Cesarman et al., 1995, Blood 86, 2708-2714), TS and DHFR (which is present in HVS, see Albrecht et al., 1992, J. Virol. 66, 5047-5058, but not human herpesviruses), encode proteins that are appropriate targets for existing drugs.

Molecular mimicry by KSHV of cell cycle regulatory and signaling proteins is a prominent feature of the virus. The KSHV genome has genes similar to cellular complement-binding proteins (ORF 4), cytokines (ORFs K2, K4 and K6), a bcl-2 protein (ORF 16), a cytokine transduction pathway protein (K9), an IL-8R-like protein (ORF74) and a D-type cyclin (ORF72). Additional regions coding for proteins with some similarity to MIP and IRF-like proteins are also present in the KSHV genome. There is a striking parallel between the KSHV genes that are similar to cellular genes and the cellular genes known to be induced by EBV infection. Cellular cyclin D, CD21/CR2, bcl-2, an IL-8R-like protein (EBI1), IL-6

and adhesion molecules are upregulated by EBV infection (Birkenbach et al., 1993, *J. Virol.* 67, 2209-2220; Falmero et al., 1993, *Oncogene* 8, 1049-1054; Finke et al., 1993, *Blood* 81, 459-469; 5 Finke et al., 1994, *Leukemia & Lymphoma* 17, 413-419; Jones et al., 1995, *J. Exper. Med.* 182, 1213-1221). This suggests that KSHV modifies the same signaling and regulation pathways that EBV modifies after infection, but does so by introducing exogenous genes 10 from its own genome.

Cellular defense against virus infection commonly involves cell cycle shutdown, apoptosis (for review, see Shen and Shenk, 1995, *Curr. Opin. Genet. Devel.* 5, 105-111) and elaboration of cell-mediated immunity 15 (CMI). The KSHV-encoded v-bcl-2, v-cyclin and v-IL-6 are active in preventing either apoptosis or cell cycle shutdown (Chang et al., 1996, *Nature* 382, 410). At least one of the  $\beta$ -chemokine KSHV gene products, 20 v-MIP-1, prevents CCR5-mediated HIV infection of transfected cells.  $\beta$ -chemokines are not known to be required for successful EBV infection of cells although EBV-infected B cells express higher levels of MIP-1 $\alpha$  than normal tonsillar lymphocytes (Harris et 25 al., 1993, 151, 5975-5983). The autocrine dependence of EBV-infected B cells on small and uncharacterized protein factors in addition to IL-6 (Tosato et al., 1990, *J. Virol.* 64, 3033-3041) leads to speculation that  $\beta$ -chemokines may also play a role in the EBV life 30 cycle.

KSHV has not formally been shown to be a transforming virus and genes similar to the major transforming genes of HVS and EBV are not present in the BC-1 35 strain KSHV. Nonetheless, dysregulation of cell proliferation control caused by the identified KSHV-encoded proto-oncogenes and cytokines may

contribute to neoplastic expansion of virus-infected cells. Preliminary studies suggest that subgenomic KSHV fragments can transform NIH 3T3 cells. If KSHV replication, like that of EBV, involves recombination of TR units (Raab-Traub and Flynn, 1986, Cell 47, 883-889), a monomorphic TR hybridization pattern present in a KS lesion would indicate a clonal virus population in the tumor. This is consistent with KS being a true neoplastic proliferation arising from single transformed, KS-infected cell rather than KSHV being a "passenger virus". Identification of KSHV genes similar to known oncoproteins and cell proliferation factors in the current study provides evidence that KSHV is likely to be a transforming virus.

EXPERIMENTAL DETAILS SECTION II:MOLECULAR MIMICRY OF HUMAN CYTOKINE AND CYTOKINE  
RESPONSE PATHWAY GENES BY KSHV

5 Four virus genes encoding proteins similar to two  
human macrophage inflammatory protein (MIP)  
chemokines, an IL-6 and an interferon regulatory  
factor (IRF or ICSP) polypeptide are present in the  
10 genome of Kaposi's sarcoma-associated herpesvirus  
(KSHV). Expression of these genes is inducible in  
infected cell lines by phorbol esters. vIL-6 is  
functionally active in B9 cell proliferation assays.  
It is primarily expressed in KSHV-infected  
15 hematopoietic cells rather than KS lesions. vMIP-I  
inhibits replication of CCR5-dependent HIV-1 strains  
in vitro indicating that it is functional and could  
contribute to interactions between these two viruses.  
Mimicry of cell signaling proteins by KSHV may  
20 abrogate host cell defenses and contribute to  
KSHV-associated neoplasia.

Kaposi's sarcoma-associated herpesvirus (KSHV) is a  
25 gammaherpesvirus related to Epstein-Barr virus (EBV)  
and herpesvirus saimiri (HVS). It is present in  
nearly all KS lesions including the various types of  
HIV-related and HIV-unrelated KS (Chang et al., 1994,  
*Science* 265, 1865-1869; Boshoff et al., 1995, *Lancet*  
30 345, 1043-1044; Dupin et al., 1995, *Lancet* 345,  
761-762; Schalling et al., 1995, *Nature Med.* 1,  
707-708). Viral DNA preferentially localizes to KS  
tumors (Boshoff et al., 1995, *Nature Med.* 1,  
1274-1278) and serologic studies show that KSHV is  
35 specifically associated with KS. Related  
lymphoproliferative disorders frequently occurring in  
patients with KS, such as primary effusion lymphomas

(PEL), a rare B cell lymphoma, and some forms of Castleman's disease are also associated with KSHV infection (Cesarman et al., 1995, *New Eng. J. Med.* 332, 1186-1191; Soulier et al., 1995, *Blood* 86, 1276-1280). Three KSHV-encoded cytokine-like polypeptides and a polypeptide similar to interferon regulatory factor genes have now been identified. Paradoxically, while cytokine dysregulation has been proposed to cause Kaposi's sarcoma (Ensoli et al., 1994, *Nature* 371, 674-680; Miles, 1992, *Cancer Treatment & Research* 63, 129-140), in vitro spindle cell lines used for these studies over the past decade are uniformly uninfected with KSHV (Ambroziak et al., *Science* 268, 582-583; Lebbé et al., 1995, *Lancet* 345, 1180).

To identify unique genes in the KSHV genome, genomic sequencing (see METHODS) was performed using Supercos-1 and Lambda FIX II genomic libraries from BC-1, a nonHodgkin's lymphoma cell line stably infected with both KSHV and EBV (Cesarman et al., 1995, *Blood* 86, 2706-2714). The KSHV DNA fragments KS330Bam and KS631Bam (Chang et al., 1994, *Science* 265, 1865-1869) were used as hybridization starting points for mapping and bi-directional sequencing. Open reading frame (ORF) analysis (see METHODS) of the Z6 cosmid sequence identified two separate coding regions (ORFs K4 and K6) with sequence similarity to  $\beta$ -chemokines and a third coding region (ORF K2) similar to human interleukin-6 (huIL-6); a fourth coding region (ORF K9) is present in the Z8 cosmid insert sequence with sequence similarity to interferon regulatory factor (IRF) polypeptides (Figures 3A-3C). None of these KSHV genes are similar to other known viral genes. Parenthetically, a protein with conserved cysteine motifs similar to  $\beta$ -chemokine motif signatures has recently been reported in the molluscum

contagiosum virus (MCV) genome. Neither vMIP-I nor vMIP-II has significant similarity to the MCV protein.

The cellular counterparts to these four viral genes  
5 encode polypeptides involved in cell responses to  
infection. For example, the MIP/RANTES (macrophage  
inflammatory protein/regulated on activation, normal  
T cell expressed and secreted) family of 8-10 kDa  
 $\beta$ -chemoattractant cytokines (chemokines) play an  
10 important role in virus infection-mediated  
inflammation (Cook et al., 1995, Science 269,  
1583-1585).  $\beta$ -chemokines are the natural ligand for  
CCR5 and can block entry of non-syncytium inducing  
(NSI), primary lymphocyte and macrophage-tropic HIV-1  
15 strains in vitro by binding to this HIV co-receptor  
(Cocchi et al., 1995, Science 270, 1811-1815). IL-6,  
initially described by its effect on B cell  
differentiation (Hirano et al., 1985, Proc Natl Acad  
Sci, USA 85, 5490; Kishimoto et al., 1995, Blood 86,  
20 1243-1254), has pleiotropic effects on a wide variety  
of cells and may play a pathogenic role in multiple  
myeloma, multicentric Castleman's disease (a  
KSHV-related disorder), AIDS-KS and EBV-related  
postransplant lymphoproliferative disease (Klein et  
25 al., 1995, Blood 85, 863-872; Hilbert et al., 1995, J  
Exp Med 182, 243-248; Brandt et al., 1990, Curr Topic  
Microbiol Immunol 166, 37-41; Leger et al., 1991,  
Blood 76, 2923-2930; Burger et al., 1994, Annals  
Hematol 69, 25-31; Tosato et al., 1993, J Clin Invest  
30 91, 2806-2814). IL-6 production is induced by either  
EBV or CMV infection and is an autocrine factor for  
EBV-infected lymphoblastoid cells that enhances their  
tumorigenicity in nude mice (Tosato et al., 1990, J  
Virol 64, 3033-3041; Scala et al., 1990, J Exp Med  
35 172, 61-68; Almeida et al., 1994, Blood 83, 370-376).  
Cell lines derived from KS lesions, although not  
infected with KSHV, also produce and respond to IL-6

(Miles et al., 1990, *Proc Natl Acad Sci USA* 87, 4068-4072; Yang et al., 1994, *J Immunol* 152, 943-955). While MIP and IL-6 are secreted cytokines, the IRF family of polypeptides regulate interferon-inducible genes in response to  $\gamma$ - or  $\alpha$ -/ $\beta$ -interferon cytokines by binding to specific interferon consensus sequences (ICS) within interferon-inducible promoter regions. A broad array of cellular responses to interferons is modulated by the repressor or transactivator functions of IRF polypeptides and several members (IRF-1 and IRF-2) have opposing anti-oncogenic and oncogenic activities (Sharf et al., 1995, *J Biol Chem* 270, 13063-13069; Harada et al., 1993, *Science* 259, 971-974; Weisz et al., 1994, *Internat Immunol* 6, 1125-1131; Weisz et al., 1992, *J Biol Chem* 267, 25589-25596).

The 289 bp ORF K6 (ORF MIP1) gene encodes a 10.5 kDa polypeptide (vMIP-I; MIP1) having 37.9% amino acid identity (71% similarity) to huMIP-1 $\alpha$  and slightly lower similarity to other  $\beta$ -chemokines (Figure 3A). ORF K4 also encodes a predicted 10.5 kDa polypeptide (vMIP-II; vMIP1 $\alpha$ -II) with close similarity and amino acid hydrophobicity profile to vMIP-I. The two KSHV-encoded MIP  $\beta$ -chemokines are separated from each other on the KSHV genome by 5.5 kb of intervening sequence containing at least 4 ORFs (see METHODS). Both polypeptides have conserved  $\beta$ -chemokine motifs (Figure 3A, residues 17-55) which include a characteristic C-C dicysteine dimer (Figure 3A, residues 36-37), and have near sequence identity to human MIP-1 $\alpha$  at residues 56-84. However, the two polypeptides show only 49.0% amino acid identity to each other and are markedly divergent at the nucleotide level indicating that this duplication is not a cloning artifact. The two viral polypeptides are more closely related to each other



phylogenetically than to huMIP-1a, huMIP-1b or huRANTES suggesting that they arose by gene duplication rather than independent acquisition from the host genome (see Sequence alignment in METHODS).  
5 The reason for this double gene dosage in the viral genome is unknown.

The KSHV ORF K2 (Figure 3B) encodes a hypothetical 204 residue, 23.4 kDa IL-6-like polypeptide with a  
10 hydrophobic 19 amino acid secretory signaling peptide having 24.8% amino acid identity and 62.2% similarity to the human polypeptide. vIL-6 also has a conserved sequence characteristic for IL-6-like interleukins (amino acids 101-125 of the gapped polypeptide) as  
15 well as conserved four cysteines which are present in IL-6 polypeptides (gapped alignment residue positions 72, 78, 101 and 111 in Figure 3B). IL-6 is a glycosylated cytokine and potential N-linked glycosylation sites in the vIL-6 sequence are present  
20 at gapped positions 96 and 107 in Figure 3C. The 449 residue KSHV vIRF polypeptide encoded by ORF K3 has lower overall amino acid identity (approximately 13%) to its human cellular counterparts than either of the vMIPs or the vIL-6, but has a conserved region derived  
25 from the IRF family of polypeptides (Figure 3C, gapped residues 88-121). This region includes the tryptophan-rich IRF ICS DNA binding domain although only two of four tryptophans thought to be involved in DNA binding are positionally conserved. It is  
30 preceded by an 87-residue hydrophilic N-terminus with little apparent IRF similarity. A low degree of amino acid similarity is present at the C-terminus corresponding to the IRF family transactivator/repressor region.

35

The four KSHV cell signaling pathway genes show similar patterns of expression in virus-infected

lymphocyte cell lines by Northern blotting (see METHODS). Whole RNA was extracted from BCP-1 (a cell line infected with KSHV alone) and BC-1 (EBV and KSHV coinfecting, see Cesarman et al., 1995, Blood 86, 2708-2714) with or without pretreatment with 30 ng/ml 12-O-tetradecanoylphorbol-13-acetate (TPA, Sigma, St. Louis MO) for 48 hours. While constitutive expression of these genes was variable between the two cell lines, expression of all four gene transcripts increased in BCP-1 and BC-1 cells after TPA induction (Figures 4A-4D). This pattern is consistent with expression occurring primarily during lytic phase virus replication. Examination of viral terminal repeat sequences of BCP-1 and BC-1 demonstrates that low level of virus lytic replication occurs in BCP-1 but not BC-1 without TPA induction (see METHODS), and both cell lines can be induced to express lytic phase genes by TPA treatment despite repression of DNA replication in BC-1. Lower level latent expression is also likely, particularly for vIL-6 (Figure 4C) and vIRF (Figure 4D), since these transcripts are detectable without TPA induction in BC-1 cells which are under tight latency control. To determine if in vitro KS spindle cell cultures retain defective or partial virus sequences that include these genes, DNA was extracted from four KS spindle cell lines (KS-2, KS-10, KS-13 and KS-22) and PCR amplified for vMIP-I, vMIP-II, vIL-6 and vIRF sequences (see METHODS). None of the spindle cell DNA samples were positive for any of the four genes.

vIL-6 was examined in more detail using bioassays and antibody localization studies to determine whether it is functionally conserved. Recombinant vIL-6 (rvIL-6) is specifically recognized by antipeptide antibodies which do not cross-react with huIL-6 (Figures 5A-5B) (see METHODS). vIL-6 is produced constitutively in

BCP-1 cells and increases markedly after 48 hour TPA induction, consistent with Northern hybridization experiments. The BC-1 cell line coinfectd with both KSHV and EBV only shows vIL-6 polypeptide expression after TPA induction (Figure 5A, lanes 3-4) and control EBV-infected P3HR1 cells are negative for vIL-6 expression (Figure 5A, lanes 5-6). Multiple high molecular weight bands present after TPA induction (21-25 kDa) may represent precursor forms of the polypeptide. Despite regions of sequence dissimilarity between huIL-6 and vIL-6, the virus interleukin 6 has biologic activity in functional bioassays using the IL-6-dependent mouse plasmacytoma cell line B9 (see METHODS). COS7 supernatants from the forward construct (rvIL-6) support B9 cell proliferation measured by <sup>3</sup>H-thymidine uptake indicating that vIL-6 can substitute for cellular IL-6 in preventing B9 apoptosis (Figure 6). vIL-6 supported B9 proliferation is dose dependent with the unconcentrated supernatant from the experiment shown in Figure 6 having biologic activity equivalent to approximately 20 pg per ml huIL-6.

Forty-three percent of noninduced BCP-1 cells (Figure 7A) have intracellular cytoplasmic vIL-6 immunostaining (see METHODS) suggestive of constitutive virus polypeptide expression in cultured infected cells, whereas no specific immunoreactive staining is present in uninfected control P3HR1 cells (Figure 7B). vIL-6 production was rarely detected in KS tissues and only one of eight KS lesions examined showed clear, specific vIL-6 immunostaining in less than 2% of cells (Figure 7C). The specificity of this low positivity rate was confirmed using preimmune sera and neutralization with excess vIL-6 peptides. Rare vIL-6-producing cells in the KS lesion are positive for either CD34, an endothelial cell marker (Figure

- 8A), or CD45, a pan-hematopoietic cell marker (Figure 8B), demonstrating that both endothelial and hematopoietic cells in KS lesions produce vIL-6. It is possible that these rare vIL-6 positive cells are entering lytic phase replication which has been shown to occur using the KSHV T1.1 lytic phase RNA probe. In contrast, well over half (65%) of ascitic lymphoma cells pelleted from an HIV-negative PEL are strongly positive for vIL-6 (Figure 7E) and express the plasma cell marker EMA (Cesarman et al., 1995, Blood 86, 2708-2714) indicating that either most PEL cells in vivo are replicating a lytic form of KSHV or that latently infected PEL cells can express high levels of vIL-6. No specific staining occurred with any control tissues examined including normal skin, tonsillar tissue, multiple myeloma or angiosarcoma using either preimmune or post-immune rabbit anti-vIL-6 antibody (Figure 7E and 7F).
- Virus dissemination to nonKS tissues was found by examining a lymph node from a patient with AIDS-KS who did not develop PEL. Numerous vIL-6-staining hematopoietic cells were present in this lymph node (Figure 8C) which was free of KS microscopically. vIL-6 positive lymph node cells were present in relatively B-cell rich areas and some express CD20 B cell surface antigen (Figure 8D), but not EMA surface antigen (unlike PEL cells) (Cesarman et al., 1995, Blood 86, 2708-2714). No colocalization of vIL-6 positivity with the T cell surface antigen CD3 or the macrophage antigen CD68 was detected, although phagocytosis of vIL-6 immunopositive cells by macrophages was frequently observed.
- To investigate whether the vMIP-I can inhibit NSI HIV-1 virus entry, human CD4+ cat kidney cells (CCC/CD4) were transiently transfected with plasmids

expressing human CCR5 and vMIP-I or its reverse construct 1-PIMv (see CCR5 and vMIP-I cloning in METHODS). These cells were infected with either M23 or SF162 primary NSI HIV-1 isolates which are known to use CCR5 as a co-receptor (Clapham et al., 1991, *J Virol* 66, 3531-3537) or with the HIV-2 variant ROD/B which can infect CD4+ CCC cells without human CCR5. Virus entry and replication was assayed by immunostaining for retroviral antigen production (Figure 9). vMIP-I cotransfection reduced NSI HIV-1 foci generation to less than half that of the reverse-construct negative control but had no effect on ROD/B HIV-2 replication.

Molecular piracy of host cell genes is a newly recognized feature of some DNA viruses, particularly herpesviruses and poxviruses (Murphy, 1994, *Infect Agents Dis* 3, 137-154; Albrecht et al., 1992, *J Virol* 66, 5047-5058; Gao and Murphy, 1994, *J Biol Chem* 269, 28539-28542; Chee et al., 1990, *Curr Top Microbiol Immunol* 154, 125-169; Massung et al., 1994, *Virol* 201, 215-240). The degree to which KSHV has incorporated cellular genes into its genome is exceptional. In addition to vMIP-I and vMIP-II, vIL-6 and vIRF, KSHV also encodes polypeptides similar to bcl-2 (ORF 16), cyclin D (ORF 72), complement-binding proteins similar to CD21/CR2 (ORF 4), an NCAM-like adhesion protein (ORF K14), and an IL-8 receptor (ORF 74). EBV also either encodes (BHRF1/bcl-2) or induces (CR-2; cyclin D; IL-6; bcl-2; adhesion molecules and an IL-8R-like EBV protein) these same cellular polypeptides (Cleary et al., 1986, *Cell* 47, 19-28; Tosato et al., 1990, *J Virol* 64, 3033-3041; Palmero et al., 1993, *Oncogene* 8, 1049; Larcher et al., 1995, *Eur J Immunol* 25, 1713-1719; Birkenbach et al., 1993, *J Virol* 67, 2209-2220). Thus, both viruses may modify similar host cell signaling and regulatory pathways.

EBV appears to effect these changes through induction of cellular gene expression whereas KSHV introduces the polypeptides exogenously from its own genome.

5 Identification of these virus-encoded cellular-like polypeptides leads to speculation about their potential roles in protecting against cellular antiviral responses. huIL-6 inhibits  $\gamma$ -interferon-induced, Bax-mediated apoptosis in  
10 myeloma cell lines (Lichtenstein et al., 1995, Cellular Immunology 162, 248-255) and vIL-6 may play a similar role in infected B cells. KSHV-encoded vIRF, vbc1-2 and v-cyclin may also interfere with host-cell mediated apoptosis induced by virus  
15 infection and v-cyclin may prevent G1 cell cycle arrest of infected cells. Interference with interferon-induced MHC antigen presentation and cell-mediated immune response (Holzinger et al., 1993, Immunol Let 35, 109-117) by vIRF is also possible.  
20 The  $\beta$ -chemokine polypeptides vMIP-I and vMIP-II may have agonist or antagonist signal transduction roles. Their sequence conservation and duplicate gene dosage are indicative of a key role in KSHV replication and survival.

25 Uncontrolled cell growth from cell-signaling pathway dysregulation is an obvious potential by-product of this virus strategy. Given the paucity of vIL-6 expressing cells in KS lesions, it is unlikely that  
30 vIL-6 significantly contributes to KS cell neoplasia. KSHV induction of hu-IL6, however, with subsequent induction of vascular endothelial growth factor-mediated angiogenesis (Holzinger et al., 1993, Immunol Let 35, 109-117), is a possibility. vIL-6  
35 could also potentially contribute to the pathogenesis of KSHV-related lymphoproliferative disorders such as PEL or the plasma cell variant of Castleman's disease.

The oncogenic potential of cellular cyclin and bcl-2 overexpression is well-established and these virus-encoded polypeptides may also contribute to KSHV-related neoplasia.

5

KSHV vMIP-1 inhibits NSI HIV-1 replication in vitro (Figure 9). Studies from early in the AIDS epidemic indicate that survival is longer for AIDS-KS patients than for other AIDS patients, and that 93% of US AIDS patients surviving >3 years had KS compared to only 28% of remaining AIDS patients dying within 3 years of diagnosis (Hardy, 1991, *J AIDS* 4, 386-391; Lamp et al., 1990, *J Am Med Assoc* 263, 402-406; Rothenberg et al., 1987, *New Eng J Med* 317, 1297-1302; Jacobson et al., 1993, *Am J Epidemiol* 138, 953-964; Lundgren et al., 1995, *Am J Epidemiol* 141, 652-658). This may be due to KS occurring at relatively high CD4+ counts and high mortality for other AIDS-defining conditions. Recent surveillance data also indicates that the epidemiology of AIDS-KS is changing as the AIDS epidemic progresses (*ibid*).

10

15

20

#### METHODS

25

30

Genomic Sequencing. Genomic inserts were randomly sheared, cloned into M13mp18, and sequenced to an average of 12-fold redundancy with complete bidirectional sequencing. The descriptive nomenclature of KSHV polypeptides is based on the naming system derived for herpesvirus saimiri (Albrecht et al., 1992, *J Virol* 66, 5047-5058).

35

Open reading frame (ORF) analysis. Assembled sequence contigs were analyzed using MacVector (IBI-Kodak, Rochester NY) for potential open reading frames greater than 25 amino acid residues and analyzed using BLASTX and BEAUTY-BLASTX (Altschul et al., 1990, *J Mol*

Biol 215, 403-410; Worley et al., 1995, Genome Res 5, 173-184; [http://dot.imgen.bcm.tmc.edu:9331/seq-search/nucleic\\_acid-search.html](http://dot.imgen.bcm.tmc.edu:9331/seq-search/nucleic_acid-search.html). Similar proteins aligned to the four KSHV polypeptides (in italics):

5 included (name (species, sequence bank accession number, smallest sum Poisson distribution probability score)): (1) vMIP-I: LD78 (MIP-1 $\alpha$ ) (human, gi 127077, p=9.8xe-22), MIP-1 $\alpha$  (Rattus, gi 790633, p=3.3xe-20), MIP-1 $\alpha$  (Mus, gi 127079, p=1.7xe-19), MIP-1 $\beta$  (Mus, gi 1346534, p=7.8xe-18); (2) vMIP-II: LD78 (MIP-1 $\alpha$ ) (human, gi 127077, p=7.1xe-23), MIP-1 $\alpha$  (Mus, gi 127079, p=8.9xe-21), MIP-1 $\alpha$  (Rattus, gi 790633, p=1.2xe-20), MIP-1 $\beta$  (Mus, gi 1346534, p=3.8xe-20); (3)

10 vIL-6: 26 kDa polypeptide (IL-6) (human, gi 13835, p=7.2xe-17), IL-6 (Macaca, gi 514386, p=1.6xe-16); and

15 (4) vIRF: ICSEB (Gallus, gi662355, p=1.1xe-11), ICSEB (Mus, sp p23611, p=1.0xe-10), lymphoid specific interferon regulatory factor (Mus, gi 972949, p=2.0xe-10), ISGF3 (Mus, gi 1263310, p=6.1xe-10), IRF4 (human, gi 1272477, p=1.0xe-9), ISGF3 (human, sp Q00978, 3.9xe-9), ICSEB (human, sp Q02556, p=2.3xe-8).

20

Sequence alignment. Amino acid sequences were aligned using CLUSTAL W (Thompson et al., 1994, Nuc Acids Res 22, 4673-4680) and compared using PAUP 3.1.1. Both

25 rooted and unrooted bootstrap comparisons produced phylogenetic trees having all 100 bootstrap replicates with viral polypeptides being less divergent from each other than from the human polypeptides.

30 Northern blotting. Northern blotting was performed using standard conditions with random-labeled probes (Chang et al., 1994, Science 265, 1865-1869) derived from PCR products for the following primer sets:

35 vMIP-I: 5'-AGC ATA TAA GGA ACT CGG CGT TAC-3' (SEQ ID NO:4), 5'-GGT AGA TAA ATC CCC CCC CTT TG-3' (SEQ ID NO:5); vMIP-II: 5'-TGC ATC AGC TTC TTC ACC CAG-3' (SEQ



ID NO:6), 5'-TGC TGT CTC GGT TAC CAG AAA AG-3' (SEQ ID NO:7); VIL-6: 5'-TCA CGT CGC TCT TTA CTT ATC GTG-3' (SEQ ID NO:8), 5'-CGC CCT TCA GTG AGA CTT CGT AAC-3' (SEQ ID NO:9); VIRF: 5'-CTT GCG ATG AAC CAT CCA GG-3' (SEQ ID NO:10); 5'-ACA ACA CCC AAT TCC CCG TC-3' (SEQ ID NO:11) on total cell RNA extracted with RNazol according to manufacturer's instructions (TelTest Inc, Friendswood TX) and 10 µg of total RNA was loaded in each lane. BCP-1, BC-1 and P3HR1 were maintained in culture conditions and induced with TPA as previously described (Gao et al., 1996, *New Eng J Med* 335, 233-241). PCR amplification for these viral genes was performed using the VMIP-I, VMIP-II, VIL-6, and VIRF primer sets with 35 amplification cycles and compared to dilutions of whole BC-1 DNA as a positive control using PCR conditions previously described (Moore and Chang, 1995, *New Eng J Med* 332, 1181-1185). KS spindle cell line DNA used for these experiments was described in Dictor et al., 1996, *Am J Pathol* 148, 2009-2016. Amplifiability of DNA samples was confirmed using human HLA-DQ alpha and pyruvate dehydrogenase primers.

VIL-6 cloning. VIL-6 was cloned from a 695 bp polymerase chain reaction (PCR) product using the following primer set: 5'-TCA CGT CGC TCT TTA CTT ATC GTG-3' (SEQ ID NO:12) and 5'-CGC CCT TCA GTG AGA CTT CGT AAC-3' (SEQ ID NO:13), amplified for 35 cycles using the 0.1 µg of BC-1 DNA as a template. PCR product was initially cloned into pCR 2.1 (Invitrogen, San Diego CA) and an EcoRV insert was then cloned into the pMET7 expression vector (Takebe et al., 1988, *Mol Cell Biol* 8, 466-472) and transfected using DEAE-dextran with chloroquine into COS7 cells (CRL-1651, American Type Culture Collection, Rockville MD). The sequence was also cloned into the pMET7 vector in the reverse orientation (6-LIV) relative to

the SRA promoter as a negative control. With orientation and sequence fidelity of both constructs confirmed by bidirectional sequencing using dye-primer chemistry on an ABI 377 sequencer (Applied Biosystems Inc, Foster City CA).

15 ml of serum-free COS7 supernatants were concentrated to 1.5 ml by ultrafiltration with a Centrplus 10 filter (Amicon, Beverly MA) and 100  $\mu$ l of supernatant concentrate or 1  $\mu$ g of rhIL-6 (R&D Systems, Minneapolis MN) was loaded per each lane in Laemmli buffer. For cell lysate immunoblotting, exponential phase cells with and without 20 ng/ml TPA induction for 48 hours were pelleted and 100  $\mu$ g of whole cell protein solubilized in Laemmli buffer was loaded per lane, electrophoresed on a 15% SDS-polyacrylamide gel and immunoblotted and developed using standard conditions (Gao et al., 1996, New Eng J Med 335, 233-241) with either rabbit anti-peptide antibody (1:100-1:1000 dilution) or anti-huIL-6 (1  $\mu$ g per ml, R&D Systems, Minneapolis MN).

Cell line B9. B9 mouse plasmacytoma cell line were maintained in Iscove's Modified Dulbecco's Medium (IMDM) (Gibco, Gaithersburg, MD), 10% fetal calf serum, 1% penicillin/streptomycin, 1% glutamine, 50  $\mu$ M  $\beta$ -mercaptoethanol, and 10 ng per ml rhIL-6 (R&D Systems, Minneapolis, MN).  $^3$ H-thymidine uptake was used to measure B9 proliferation in response to huIL-6 or recombinant supernatants according to standard protocols (R&D Systems, Minneapolis, MN). Briefly, serial 1:3 dilutions of huIL-6 or Centrplus 10 concentrated recombinant supernatants were incubated with  $2 \times 10^4$  cells per well in a 96 well plate for 24 hours at 37°C with 10  $\mu$ l of thymidine stock solution (50  $\mu$ l of 1mCi/ml  $^3$ H-thymidine in 1 ml IMDM) added to each well during the final four hours of incubation.

Cells were harvested and incorporated  $^3\text{H}$ -thymidine determined using a liquid scintillation counter. Each data point is the average of six determinations with standard deviations shown.

5  
vIL-6 immunostaining. Immunostaining was performed using avidin-biotin complex (ABC) method after deparaffinization of tissues and quenching for 30 minutes with 0.03%  $\text{H}_2\text{O}_2$  in PBS. The primary antibody  
10 was applied at a dilution of 1:1250 after blocking with 10% normal goat serum, 1% BSA, 0.5% Tween 20. The secondary biotinylated goat anti-rabbit antibody (1:200 in PBS) was applied for 30 minutes at room temperature followed by three 5 minute washes in PBS.  
15 Peroxidase-linked ABC (1:100 in PBS) was applied for 30 minutes followed by three 5 minute washes in PBS. A diaminobenzidine (DAB) chromogen detection solution (0.25% DAB, 0.01%  $\text{H}_2\text{O}_2$  in PBS) was applied for 5 minutes. Slides are then washed, counterstained with  
20 hematoxylin and coverslipped. Amino ethyl carbazole (AEC) or Vector Red staining was also used allowing better discrimination of double-labeled cells with Fast Blue counterstaining for some surface antigens. For CD68, in which staining might be obscured by vIL-6  
25 cytoplasmic staining, double label immunofluorescence was used. Microwaved tissue sections were blocked with 2% human serum, 1% bovine serum albumin (BSA) in PBS for 30 minutes, incubated overnight with primary antibodies and developed with fluorescein-conjugated  
30 goat anti-rabbit IgG (1:100, Sigma) for vIL-6 localization and rhodamine-conjugated horse anti-mouse IgG (1:100, Sigma) for CD68 localization for 30 minutes. After washing, secondary antibody incubation was repeated twice with washing for 15 minutes each to  
35 amplify staining. For the remaining membrane antigens, slides were developed first for vIL-6 and then then secondly with the cellular antigen, as well

as the reverse localization (cellular antigen antibody first, anti-VIL-6 second) to achieve optimal visualization and discrimination of both antigens. In each case, the first antibody was developed using AEC (Sigma) with blocking solution preincubation (1% BSA, 10% normal horse serum, 0.5% Tween 20 for 30 minutes) and development per manufacturer's instructions. The second antibody was developed using the ABC-alkaline phosphatase technique with Fast Blue chromagen. Both microwaving and trypsinization resulted in poorer localization and specificity of VIL-6 immunolocalization. In cases where this was required for optimal localization of membrane antigen, these techniques were applied after VIL-6 AEC localization. Vector-Red (Vector, Burlingame, CA) staining was used as an alternative stain to AEC to achieve optimal discrimination and was performed per manufacturer's protocol using the ABC-alkaline phosphatase technique. Cell antigen antibodies examined included CD68 (1:800, from clone Kim 6), epithelial membrane antigen (EMA, 1:500, Dako, Carpinteria, CA), CD3 (1:200, Dako), CD20 (1:200, Dako), OPD4 (1:100, Dako), CD34 (1:15, Dako), CD45 (1:400, from clone 9.4), L26 (1:100, Immunotech, Westbrook, ME) and Leu23 (1:100, Becton-Dickinson, San Jose, CA) on tissues prepared according to manufacturer's instructions. Specific VIL-6 colocalization was only found with CD34 and CD45 in KS lesions, EMA in PEL, and CD20 and CD45 in lymph node tissues.

Immunohistochemical VIL-6 localization was performed on exponential phase BCP-1 cells with or without 48 hour TPA incubation after embedding in 1% agar in saline. The percentages of positive cells were determined from cell counts of three random high power microscopic fields per slide. Lower percentages of BCP-1 cells stain positively for VIL-6 after TPA

treatment possibly reflecting cell lysis and death from lytic virus replication induction by TPA. Immunostaining of cells and tissues was demonstrated to be specific by neutralization using overnight incubation of antisera with 0.1 µg/ml vIL-6 synthetic peptides at 4°C and by use of preimmune rabbit antisera run in parallel with the postimmune sera for the tissues or cell preparations. No specific staining was seen after either peptide neutralization or use of preimmune sera.

CCR5 and vMIP-I cloning. CCR5 was cloned into pRCMV vector (Invitrogen) and both forward and reverse orientations of the vMIP-I gene were cloned into pMET7 after PCR amplification using the following primer pairs: 5'-AGC ATA TAA GGA ACT CGG CGT TAC-3' (SEQ ID NO:14), 5'-GGT AGA TAA ACT CCC CCC CTT TG-3' (SEQ ID NO:15). CCR5 alone and with the forward construct (vMIP-I), the reverse construct (I-PIMV) and empty pMET7 vector were transfected into CCC/CD4 cells (CCC cat cells stably expressing human CD4, see McKnight et al., 1994, Virol 201, 8-18) using Lipofectamine (Gibco). After 48 hours, media was removed from the transfected cells and 1000 TCID<sub>50</sub> of SF162, M23 or ROD/B virus culture stock was added. Cells were washed four times after 4 hours of virus incubation and grown in DMEM with 5% FCS for 72 hours before immunostaining for HIV-1 p24 or HIV-2 gp105 as previously described. Each condition was replicated 3-4 times (Figure 9) with medians and error bars representing the standard deviations expressed as percentages of the CCR5 alone foci.

EXPERIMENTAL DETAILS SECTION III:

The following patents are hereby incorporated by reference to more fully describe the invention described herein:

1. Fowlkes, CARBOXY TERMINAL IL-6 MUTEINS, PATENT NO. 5,565,336, ISSUED October 15, 1996;
2. Skelly et al., METHOD OF MAKING CYSTEINE DEPLETED IL-6 MUTEINS, PATENT NO. 5,545,537, ISSUED August 13, 1996;
3. Ulrich, COMPOSITION AND METHOD FOR TREATING INFLAMMATION, PATENT NO. 5,376,366, ISSUED December 27, 1994;
4. Skelly et al., CYSTEINE DEPLETED IL-6 MUTEINS, PATENT NO. 5,359,034, ISSUED October 25, 1994;
5. Williams, ULTRAPURE HUMAN INTERLEUKIN 6, PATENT NO. 5,338,834, ISSUED August 16, 1994;
6. Fowlkes, CARBOXY TERMINAL IL-6 MUTEINS, PATENT NO. 5,338,833, ISSUED August 16, 1994;
7. Ulrich, COMPOSITION AND METHOD FOR TREATING INFLAMMATION, PATENT NO. 5,300,292, ISSUED April 05, 1994;
8. Mikayama et al., MODIFIED HIL-6, PATENT NO. 5,264,209, ISSUED November 23, 1993;
9. Park, HYPERGLYCOSYLATED CYTOKINE CONJUGATES, PATENT NO. 5,217,881, ISSUED June 08, 1993;

10. Goldberg and Faquin, INTERLEUKIN 6 TO STIMULATE  
ERYTHROPOIETIN PRODUCTION, PATENT NO. 5,188,828,  
ISSUED February 23, 1993;
- 5 11. Miles et al., METHOD TO TREAT KAPOSI'S SARCOMA,  
PATENT NO. 5,470,824, ISSUED November 28, 1995;
12. Li and Ruben, MACROPHAGE INFLAMMATORY PROTEIN  
10 -3 AND -4 [Isolated polynucleotide encoding said  
polypeptide], PATENT NO. 5,504,003, ISSUED April 22,  
1996;
13. Gewirtz, SUPPRESSION OF MEGAKARYOCYTOPOIESIS BY  
15 MACROPHAGE INFLAMMATORY PROTEINS [Reducing number of  
circulating platelets in bloodstream], PATENT NO.  
5,306,709, ISSUED April 26, 1994;
14. Fahey et al., METHOD AND AGENTS FOR PROMOTING  
20 WOUND HEALING, PATENT NO. 5,145,676, ISSUED September  
8, 1992;
15. Rosen et al., POLYNUCLEOTIDE ENCODING MACROPHAGE  
INFLAMMATORY PROTEIN GAMMA, PATENT NO. 5,556,767,  
25 ISSUED September 17, 1996;
16. Chuntharapai et al., ANTIBODIES TO HUMAN IL-8  
TYPE A RECEPTOR, PATENT NO. 5,543,503, ISSUED August  
06, 1996;
- 30 17. Chuntharapai et al., ANTIBODIES TO HUMAN IL-8  
TYPE B RECEPTOR [A monoclonal antibody as  
antiinflammatory agent treating an inflammatory  
disorder], PATENT NO. 5,440,021, ISSUED August 08,  
35 1995;

18. Kunkel et al., LABELLED MONOCYTE CHEMOATTRACTANT PROTEIN MATERIAL AND MEDICAL USES THEREOF, PATENT NO. 5,413,778, ISSUED May 9, 1995;
- 5 19. Lyle and Kunkel, LABELLED INTERLEUKIN-8 AND MEDICAL USES THEREOF [Radionuclide labeled chemokines, imaging agents], PATENT NO. 5,346,686, ISSUED September 13, 1994;
- 10 20. Jones et al., ANTI-CANCER QUINAZOLINE DERIVATIVES, PATENT NO. 4,564,616, ISSUED January 14, 1986;
- 15 21. DeGraw et al., ANTIINFLAMMATORY AND ANTINEOPLASTIC 5-DEAZAAMINOPTERINS AND 8,10-DIDEAZAAMINOPTERINS, PATENT NO. 5,536,724, ISSUED July 16, 1996;
- 20 22. Mahan et al., IN VIVO SELECTION OF MICROBIAL VIRULENCE GENES [Genetic engineering and expression using auxotrophic or antibiotic sensitive microorganism's chromosome], PATENT NO. 5,434,066, ISSUED July 18, 1995;
- 25 23. DeGraw et al., 8,10-DIDEAZATETRAHYDROFOLIC ACID DERIVATIVES [Antitumor agents], PATENT NO. 5,167,963, ISSUED December 1, 1992; and
- 30 24. Watanabe, 6,7-DIHYDROPYRROL[3,4-C]PYRIDO[2,3-D]PYRIMIDINE DERIVATIVES [STRUCTURALLY SIMILAR TO THYMIDYLIC ACID], PATENT NO. 4,925,939, ISSUED May 15, 1990.



REFERENCES

1. Chang, Yuan, E Cesarman, MS Pessin, F Lee,  
J Culpepper, DM Knowles and Patrick S Moore  
(1994) Identification of herpesvirus-like  
DNA sequences in AIDS-associated Kaposi's  
sarcoma. *Science* 265, 1865-1869.
2. Moore, Patrick S and Yuan Chang (1995)  
Detection of herpesvirus-like DNA sequences  
in Kaposi's sarcoma in patients with and  
those without HIV infection. *New Eng J Med*  
332, 1181-1185.
3. Cesarman, E, Yuan Chang, Patrick S Moore, JW  
Said and DM Knowles (1995) Kaposi's sarcoma-  
associated herpesvirus-like DNA sequences  
are present in AIDS-related body cavity  
based lymphomas. *New Eng J Med* 332, 1186-  
1191.
4. Cesarman, E, Patrick S Moore, PH Rao, G  
Inghirami, DM Knowles and Yuan Chang (1995)  
In vitro establishment and characterization  
of two AIDS-related lymphoma cell lines  
containing Kaposi's-sarcoma associated  
herpesvirus-like (KSHV) DNA sequences.  
*Blood* 86, 2708-2714.

Table 1. KSHV Genome ORFs and their similarity to genes in other herpesviruses.

| Name   | Pol | Start  | Stop   | Size | HVS<br>%Sim | HVS<br>%Id | BBV Name | BBV<br>%Sim | BBV<br>%Id |
|--------|-----|--------|--------|------|-------------|------------|----------|-------------|------------|
| K1     | -   | 105    | 974    | 289  |             |            |          |             |            |
| ORF4*  | -   | 1142   | 2794   | 550  | 45.3        | 31.2       |          |             |            |
| **     |     |        |        |      | 46.4        | 34.0       |          |             |            |
| ORF6   | -   | 3210   | 6611   | 1133 | 74.1        | 55.2       | BALF2    | 65.6        | 42.1       |
| ORF7   | -   | 6628   | 8715   | 695  | 65.0        | 44.7       | BALF3    | 59.9        | 41.3       |
| ORF8   | -   | 8699   | 11,236 | 845  | 72.5        | 54.9       | BALF4    | 62.1        | 42.6       |
| ORF9   | -   | 11,363 | 14,401 | 1012 | 77.6        | 60.1       | BALF5    | 70.9        | 55.6       |
| ORF10  | +   | 14,519 | 15,775 | 418  | 50.4        | 26.2       |          |             |            |
| ORF11  | +   | 15,790 | 17,012 | 407  | 49.4        | 28.9       | Raji LFD | 44.4        | 27.9       |
| K2     | -   | 17,875 | 17,261 | 204  |             |            |          |             |            |
| ORF01  | -   | 18,553 | 17,921 | 210  | 65.8        | 48.4       |          |             |            |
| K3     | -   | 19,609 | 18,608 | 333  |             |            |          |             |            |
| ORF70  | -   | 21,104 | 20,091 | 337  | 79.5        | 66.4       |          |             |            |
| K4     | -   | 21,832 | 21,548 | 94   |             |            |          |             |            |
| K5     | -   | 26,483 | 25,713 | 257  |             |            |          |             |            |
| K6     | -   | 27,424 | 27,137 | 95   |             |            |          |             |            |
| K7     | -   | 28,622 | 29,000 | 126  |             |            |          |             |            |
| ORF16  | +   | 30,145 | 30,672 | 175  | 50.0        | 26.7       | BHRF1    | 46.3        | 22.8       |
| ORF17  | -   | 32,482 | 30,821 | 553  | 60.3        | 42.9       | BVRF2    | 58.8        | 34.1       |
| ORF18  | -   | 32,424 | 33,197 | 257  | 70.6        | 48.4       |          |             |            |
| ORF19  | -   | 34,842 | 33,194 | 549  | 62.8        | 43.8       | BVRF1    | 62.5        | 42.0       |
| ORF20  | -   | 35,573 | 34,611 | 320  | 59.6        | 42.7       | BXRF1    | 54.7        | 34.6       |
| ORF21  | -   | 35,383 | 37,125 | 580  | 50.9        | 32.5       | BXLF1    | 50.7        | 28.2       |
| ORF22  | -   | 37,113 | 39,305 | 730  | 53.9        | 35.2       | BXLF2    | 48.3        | 26.9       |
| ORF23  | -   | 40,316 | 39,302 | 404  | 57.4        | 33.7       | BTRF1    | 51.0        | 31.0       |
| ORF24  | -   | 42,778 | 40,520 | 752  | 63.6        | 45.6       | BCRF1    | 56.4        | 37.7       |
| ORF25  | +   | 42,775 | 46,907 | 1376 | 80.9        | 65.8       | BDLF1    | 74.8        | 56.8       |
| ORF26  | +   | 46,933 | 47,850 | 305  | 76.8        | 58.3       | BDLF1    | 73.4        | 48.8       |
| ORF27  | -   | 47,673 | 48,745 | 290  | 49.6        | 29.6       | BDLF2    | 43.3        | 19.6       |
| ORF28  | -   | 48,991 | 49,299 | 102  | 41.2        | 21.7       | BDLF3    |             |            |
| ORF29b | -   | 50,417 | 49,362 | 351  | 41.8        | 17.0       | BDRF1    | 42.3        | 16.3       |
| ORF30  | +   | 50,623 | 50,856 | 77   | 52.1        | 31.0       | BDLF3.5  |             |            |
| ORF31  | +   | 50,763 | 51,437 | 224  | 62.0        | 43.5       | BDLF4    | 58.9        | 36.4       |
| ORF32  | +   | 51,404 | 52,768 | 454  | 51.7        | 30.1       | BGLF1    | 47.0        | 26.6       |
| ORF33  | +   | 52,761 | 53,699 | 312  | 58.6        | 36.4       | BGLF2    | 52.8        | 32.2       |
| ORF29a | -   | 54,676 | 53,738 | 312  | 41.9        | 15.8       | BGRF1    | 57.1        | 40.6       |
| ORF34  | +   | 54,675 | 55,658 | 327  | 58.9        | 40.7       | BGLF3    | 54.8        | 33.0       |
| ORF35  | +   | 55,639 | 56,091 | 151  | 60.0        | 31.7       | BGLF3.5  |             |            |
| ORF36  | +   | 55,976 | 57,310 | 444  | 49.4        | 31.2       | BGLF4    | 50.0        | 30.2       |
| ORF37  | +   | 57,273 | 58,733 | 486  | 63.9        | 50.4       | BGLF5    | 60.1        | 42.7       |
| ORF38  | +   | 58,688 | 58,873 | 61   | 58.6        | 39.7       | BBLF1    | 52.5        | 23.0       |
| ORF39  | -   | 60,175 | 58,976 | 399  | 73.0        | 52.1       | BBRF3    | 65.2        | 43.6       |
| ORF40  | +   | 60,308 | 61,681 | 457  | 51.9        | 28.1       | BBLF2    | 47.1        | 23.3       |
| ORF41  | +   | 61,827 | 62,444 | 205  | 53.4        | 29.0       | BBLF3    |             |            |
| ORF42  | -   | 63,272 | 62,436 | 278  | 55.8        | 38.9       | BBRF2    | 52.9        | 33.0       |
| ORF43  | -   | 64,953 | 63,136 | 605  | 74.9        | 60.5       | BBRF1    | 67.6        | 50.1       |
| ORF44  | +   | 64,892 | 67,256 | 786  | 75.5        | 61.4       | BBLF4    | 67.8        | 51.1       |
| ORF45  | -   | 68,576 | 67,353 | 407  | 50.2        | 30.7       | BKRF4    | 48.9        | 26.2       |
| ORF46  | -   | 69,404 | 68,637 | 255  | 73.0        | 59.5       | BKRF3    | 69.2        | 54.8       |
| ORF47  | -   | 69,915 | 69,412 | 167  | 52.0        | 29.9       | BKRF4    | 53.8        | 24.2       |
| ORF48  | -   | 71,381 | 70,173 | 402  | 47.3        | 24.4       | BBRF2    | 46.1        | 18.8       |
| ORF49  | -   | 72,538 | 71,630 | 302  | 45.4        | 21.2       | BBRF1    | 49.8        | 28.0       |
| ORF50  | +   | 72,734 | 74,625 | 631  | 46.5        | 24.9       | BRLF1    | 41.4        | 19.0       |
| K8     | +   | 74,850 | 75,589 | 239  |             |            |          |             |            |
| ORF52  | -   | 77,197 | 76,802 | 121  | 50.0        | 33.3       | BLRF2    | 54.6        | 36.9       |
| ORF53  | -   | 77,665 | 77,333 | 110  | 59.6        | 36.0       | BLRF1    | 58.1        | 40.9       |
| ORF54  | +   | 77,667 | 78,623 | 318  | 55.0        | 35.5       | ELLF3    | 53.7        | 32.4       |
| ORF55  | -   | 78,448 | 76,765 | 227  | 64.4        | 46.4       | BSRF1    | 61.6        | 44.0       |
| ORF56  | -   | 79,436 | 81,967 | 843  | 62.5        | 44.3       | BSLF1    | 56.6        | 35.4       |

126

|       |   |         |         |      |      |      |       |      |      |
|-------|---|---------|---------|------|------|------|-------|------|------|
| ORF57 | + | 82,717  | 83,544  | 275  | 56.9 | 31.5 | BMLF1 | 45.1 | 22.1 |
| K9    | - | 85,109  | 83,860  | 449  |      |      |       |      |      |
| K10   | - | 88,164  | 86,074  | 696  |      |      |       |      |      |
| K11   | - | 93,367  | 91,964  | 467  |      |      |       |      |      |
| ORF58 | - | 95,544  | 94,471  | 257  | 55.9 | 28.7 | BMRF2 | 50.6 | 25.3 |
| ORF59 | - | 96,729  | 95,549  | 396  | 54.1 | 32.3 | BMRF1 | 50.7 | 28.3 |
| ORF60 | - | 97,787  | 96,870  | 305  | 79.3 | 64.6 | BaRF1 | 74.8 | 57.5 |
| ORF61 | - | 100,194 | 97,816  | 792  | 69.4 | 52.4 | SORF2 | 64.1 | 43.6 |
| ORF62 | - | 101,194 | 100,199 | 331  | 64.6 | 40.2 | BORF1 | 57.7 | 34.7 |
| ORF63 | - | 101,208 | 103,994 | 927  | 52.1 | 32.1 | BOLF1 | 47.0 | 24.5 |
| ORF64 | - | 104,000 | 111,907 | 2635 | 50.1 | 29.7 | BPLF1 | 46.6 | 26.1 |
| ORF65 | - | 112,443 | 111,931 | 170  | 60.4 | 40.3 | BFRF3 | 49.4 | 27.8 |
| ORF66 | - | 113,759 | 112,470 | 429  | 58.7 | 34.7 | BFRF2 | 50.0 | 28.0 |
| ORF67 | - | 114,508 | 113,693 | 271  | 71.8 | 53.0 | BFRF1 | 62.8 | 39.5 |
| ORF68 | + | 114,768 | 116,405 | 545  | 64.7 | 45.4 | BFLF1 | 58.3 | 36.1 |
| ORF69 | - | 116,669 | 117,346 | 225  | 71.1 | 53.6 | BFLF2 | 60.7 | 41.7 |
| K12   | - | 118,101 | 117,919 | 60   |      |      |       |      |      |
| K13   | - | 122,710 | 122,291 | 139  |      |      |       |      |      |
| ORF72 | - | 123,566 | 122,793 | 257  | 53.0 | 32.5 |       |      |      |
| ORF73 | - | 127,296 | 123,808 | 1162 | 51.2 | 31.8 |       |      |      |
| K14   | - | 127,683 | 128,929 | 348  |      |      |       |      |      |
| ORF74 | - | 129,371 | 130,399 | 342  | 57.8 | 34.1 |       |      |      |
| ORF75 | - | 134,440 | 130,550 | 1296 | 54.8 | 36.3 | BNRF1 |      |      |
| K15   | - | 136,179 | 135,977 | 100  |      |      |       |      |      |

| Name   | Function                           |
|--------|------------------------------------|
| K1     |                                    |
| ORF4*  | Complement binding protein (v-CBP) |
| **     |                                    |
| ORF6   | ssDNA binding protein (SSBP)       |
| ORF7   | Transport protein                  |
| ORF8   | Glycoprotein B (gB)                |
| ORF9   | DNA polymerase (pol)               |
| ORF10  |                                    |
| ORF11  |                                    |
| K2     | vIL-6                              |
| ORF02  | DHFR                               |
| K3     | BHV4-IE1 I                         |
| ORF70  | Thymidylate synthase (TS)          |
| K4     | VMIP-II                            |
| K5     | BHV4-IE1 II                        |
| K6     | VMIP-I                             |
| K7     |                                    |
| ORF16  | Bcl-2                              |
| ORF17  | Capsid protein I                   |
| ORF18  |                                    |
| ORF19  | Tegument protein I                 |
| ORF20  |                                    |
| ORF21  | Thymidine kinase (TK)              |
| ORF22  | Glycoprotein H (gH)                |
| ORF23  |                                    |
| ORF24  |                                    |
| ORF25  | Major capsid protein (MCP)         |
| ORF26  | Capsid protein II                  |
| ORF27  |                                    |
| ORF28  |                                    |
| ORF29b | Packaging protein II               |
| ORF30  |                                    |
| ORF31  |                                    |
| ORF32  |                                    |

|        |                                     |
|--------|-------------------------------------|
| ORF33  |                                     |
| ORF29a | Packaging protein I                 |
| ORF34  |                                     |
| ORF35  |                                     |
| ORF36  | Viral protein kinase                |
| ORF37  | Alkaline exonuclease (AE)           |
| ORF38  |                                     |
| ORF39  | Glycoprotein M (gM)                 |
| ORF40  | Helicase-primase, subunit 1         |
| ORF41  | Helicase-primase, subunit 2         |
| ORF42  |                                     |
| ORF43  | Capsid protein III                  |
| ORF44  | Helicase-primase, subunit 3         |
| ORF45  | Virion assembly protein             |
| ORF46  | Uracil DNA glycosylase (UDG)        |
| ORF47  | Glycoprotein L (gL)                 |
| ORF48  |                                     |
| ORF49  |                                     |
| ORF50  | Transactivator (LCTP)               |
| K8     |                                     |
| ORF52  |                                     |
| ORF53  |                                     |
| ORF54  | dUTPase                             |
| ORF55  |                                     |
| ORF56  | DNA replication protein I           |
| ORF57  | Immediate-early protein II (IEP-II) |
| K9     | VIRF1 (ICSBP)                       |
| K10    |                                     |
| K11    |                                     |
| ORF58  | Phosphoprotein                      |
| ORF59  | DNA replication protein II          |
| ORF60  | Ribonucleotide reductase, small     |
| ORF61  | Ribonucleotide reductase, large     |
| ORF62  | Assembly/DNA maturation             |
| ORF63  | Tegument protein II                 |
| ORF64  | Tegument protein III                |
| ORF65  | Capsid protein IV                   |
| ORF66  |                                     |
| ORF67  | Tegument protein IV                 |
| ORF68  | Glycoprotein                        |
| ORF69  |                                     |
| K12    | Kaposin                             |
| K13    |                                     |
| ORF72  | Cyclin D                            |
| ORF73  | Immediate-early protein (IEP)       |
| K14    | OX-2 (v-adh)                        |
| ORF74  | G-protein coupled receptor          |
| ORF75  | Tegument protein/FGARAT             |
| K15    |                                     |

Legend to Table 1. Name (e.g. K1 or ORF4; refers to the KSHV ORF designation; Pol signifies polarity of the ORF within the KSHV genome; Start refers to the position of the first LUR nucleotide in the start codon; Stop refers to the position of the last LUR nucleotide in the stop codon; Size indicates the number of amino acid residues encoded by the KSHV ORF; HVS%Sim indicates the percent similarity of the indicated KSHV ORF to the corresponding ORF of

herpesvirus saimiri; HVS%Id indicates the percent identity of the indicated KSHV ORF to the corresponding ORF of herpesvirus saimiri; EBV Name indicates the EBV ORF designation; EBV%Sim indicates the percent similarity of the indicated KSHV ORF to the named Epstein-Barr virus ORF; EBV%Id indicates the percent identity of the indicated KSHV ORF to the named Epstein-Barr virus ORF. The asterisks in the KSHV Name column indicate comparison of KSHV ORF4 to HVS ORF4a (\*) and HVS ORF4b (\*\*). The entire unannotated genomic sequence is deposited in GenBank, under the accession numbers: U75698 (LUR), U75699 (terminal repeat), and U75700 (incomplete terminal repeat). The sequence of the LUR (U75698) is also set forth in its entirety in the Sequence Listing below. Specifically, the sequence of the LUR is set forth in 5' to 3' order in SEQ ID Nos:17-20. More specifically, nucleotides 1-35,100 of the LUR are set forth in SEQ ID NO:17 numbered nucleotides 1-35,100, respectively; nucleotides 35,101-70,200 of the LUR are set forth in SEQ ID NO:18 numbered nucleotides 1-35,100, respectively; nucleotides 70,201-105,300 of the LUR are set forth in SEQ ID NO:19 numbered nucleotides 1-35,100, respectively; and nucleotides 105,301-137,507 of the LUR are set forth in SEQ ID NO:20 numbered nucleotides 1-32,207, respectively.

# SEQUENCE LISTING

GENERAL INFORMATION:

(1) APPLICANT: The Trustees of Columbia University in the City of New York

(11) TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND USES THEREOF

(111) NUMBER OF SEQUENCES: 20

(24) CORRESPONDENCE ADDRESS:

RESPONDENCE ADDRESS:  
(A) ADDRESSEE: Cooper & Dunham LLP  
(B) STREET: 1185 Avenue of the Americas  
(C) CITY: New York  
(D) STATE: New York  
(E) COUNTRY: U.S.A.  
(F) ZIP: 10036

... COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

## (4) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

(U) (1) ATTORNEY-AGENT INFORMATION:

(A) NAME: White, John P.  
(B) REGISTRATION NUMBER: 28,675  
(C) REFERENCE/DOCKET NUMBER: 45185-F-FCF

(1X) TELECOMMUNICATION INFORMATION: (1X) 776 0100

(A) TELEPHONE: (212) 376-0400  
(B) TELEFAX: (212) 391-0525

(2) INFORMATION FOR SEQ ID NO:1:

## (1) SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 338 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:1:

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Phe | Pro | Phe | Val | Pro | Leu | Ser | Leu | Tyr | Val | Ala | Lys | Lys | Leu | Phe |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Arg | Ala | Arg | Gly | Phe | Arg | Phe | Cys | Gln | Lys | Pro | Gly | Val | Leu | Ala | Leu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Ala | Pro | Glu | Val | Asp | Pro | Cys | Ser | Ile | Gln | His | Glu | Val | Thr | Gly | Ala |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Glu | Thr | Pro | His | Glu | Glu | Leu | Gln | Tyr | Leu | Arg | Gln | Leu | Arg | Glu | Ile |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |

128

Leu Cys Arg Gly Ser Asp Arg Leu Asp Arg Thr Gly Ile Gly Thr Leu  
 65 70 75 80  
 Ser Leu Phe Gly Met Gln Ala Arg Tyr Ser Leu Arg Asp His Phe Pro  
 85 90 95  
 Leu Leu Thr Thr Lys Arg Val Phe Trp Arg Gly Val Val Gln Glu Leu  
 100 105 110  
 Leu Trp Phe Leu Lys Gly Ser Thr Asp Ser Arg Glu Leu Ser Arg Thr  
 115 120 125  
 Gly Val Lys Ile Trp Asp Lys Asn Gly Ser Arg Glu Phe Leu Ala Gly  
 130 135 140  
 Arg Gly Leu Ala His Arg Arg Glu Gly Asp Leu Gly Pro Val Tyr Gly  
 145 150 155 160  
 Phe Gln Trp Arg His Phe Gly Ala Ala Tyr Val Asp Ala Asp Ala Asp  
 165 170 175  
 Tyr Thr Gly Gln Gly Phe Asp Gln Leu Ser Tyr Ile Val Asp Leu Ile  
 180 185 190  
 Lys Asn Asn Pro His Asp Arg Arg Ile Ile Met Cys Ala Trp Asn Pro  
 195 200 205  
 Ala Asp Leu Ser Leu Met Ala Leu Pro Pro Cys His Leu Leu Cys Gln  
 210 215 220  
 Phe Tyr Val Ala Asp Gly Glu Leu Ser Cys Gln Leu Tyr Gln Arg Ser  
 225 230 235 240  
 Gly Asp Met Gly Leu Gly Val Pro Phe Asn Ile Ala Ser Tyr Ser Leu  
 245 250 255  
 Leu Thr Tyr Met Leu Ala His Val Thr Gly Leu Arg Pro Gly Glu Phe  
 260 265 270  
 Ile His Thr Leu Gly Asp Ala His Ile Tyr Lys Thr His Ile Glu Pro  
 275 280 285  
 Leu Arg Leu Gln Leu Thr Arg Thr Pro Arg Pro Phe Pro Arg Leu Glu  
 290 295 300  
 Ile Leu Arg Ser Val Ser Ser Met Glu Glu Phe Thr Pro Asp Asp Phe  
 305 310 315 320  
 Arg Leu Val Asp Tyr Cys Pro His Pro Thr Ile Arg Met Glu Met Ala  
 325 330 335  
 Val \*

(2) INFORMATION FOR SEQ ID NO:2:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(12) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Thr His Tyr Ser Pro Pro Lys Phe Asp Arg

129

10

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Pro Asp Val Thr Pro Asp Val His Asp Arg  
 1 2 3 4 5 6 7 8 9 10

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AGCATATAAG GAAGTCGGCG TTAC

24

## (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGTAGATAAA TCCCCCCCCCT TTG

23

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)



(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TGATCAGCT TCTTCACCCA G

21

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TGTTTCTCTCG GTTACCAGAA AAG

23

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCACGTCGCT CTTTACTTAT CGTG

24

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CGGCTTTCAG TGAGACTTCG TAAC

24

(2) INFORMATION FOR SEQ ID NO:10:

131

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTTGGGATGA ACCATCCAGG

20

(2) INFORMATION FOR SEQ ID NO:11:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ACAACACCCA ATTCCCGTC

20

(2) INFORMATION FOR SEQ ID NO:12:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 34 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TCACGTCGCT CTTTACTTAT CGTG

24

(2) INFORMATION FOR SEQ ID NO:13:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 34 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGCCCTTCAG TGAGACTTCG TAAC

24

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 24 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGCATATAAG GAATCGGGG TTAC

24

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 23 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGTAGATAAA CTCCTCCCCCT TTG

23

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 801 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGTGAACACC CCGCGCCCCG CGCCCCCCAC ACCGCGCCGC CCCTCCCCCT CCCCCCGCTC 60  
 GCCTCCCGGC GTGCGCCCA GCGCCCGGCC GGAGCCGGCC GCCCGCGGGG GGCAGGGCGC 120  
 GCGCGGCGGC TCCCTCGCGG GCGGGGGGAC GGGGGAGGGG GCGCGCGGGC CCCCCGCGGC 180

```

CGCGGCAGCTG GAGCGCGAGG GCGCGCGCGG GCGCGCAGCG GCGGCGCAGG CCGCGGGGGG 240
CGGAGCGCGCG AGCGGGGCGG GGGTACGGGG CTAGGCCACG AATAATTTTT TTTTCGGGCG 300
GCGCGCGCGAA CTTCTCTCGG CCGCGCGGTC CCGCGCGGCG GCGCGCGCGG CCGCGGGGGG 360
GTAAAACAGG GGGGGGGGGA TCGCGCGCGG GCGCGCGCGG GCGCGCGCGG GCGCGCTTGT 420
TTCTTTTTTT CCGCGCGCGG CCGCGCGCGG AGCGCGCGCG GCGCGCGCGG CCGCGCGCTG 480
CGCGGGGGGG TCGCGGGGGG GCGCGCTGTC CCGCGCGCGG CCGCGCGCGG CCGCGCGCGG 540
CGCGCGCGGA TCGCGCGGGG GCGCGCGCGG CCGCGCGCGG AGCGCGCGG GCGCGCGCGG 600
CCTCGCGCGG GCGCATGGGG CCGCGCGCGG CCGCGCGCGG CCGCGCGCGG GCGCGCGCGG 660
CGCGCGCGCG GCGCGCGGGG GAACCGCGGG AGCGAGGGAA GGGGGCGCGG TCTCTCTACT 720
GTGCGAGGAG TCTGGGCTGC TGTGTGTGAG CCGTTTTGGG GGAGCGCTCT CAGTGTCTTG 780
TACGTGGAGC CCGGACACT A 800

```

## (3) INFORMATION FOR SEQ ID NO:17:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35100 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(2) MOLECULE TYPE: DNA (genomic)

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```

TACTAATTTT CAAAGGCGGG GTTCTGCCAG GCATAGTCTT TTTTCTTGGC GCGCCTTGTG 60
TAAACCTGTC TTTGAGACCT TGTGGACAT CCGTACAAAT CAAGATGTTG CTGTATGTTG 120
TCTGCAGTCT GCGGTTTTGC TTTGAGGAC TATTAAGCCT TTCTCTGCTA TCGTTTCCAA 180
ATTTGTGCCC TGGAGTGATT TCAACGCCTT ACACGTTGAC CTCTCTGTCT AATGCATCCT 240
TGCCAATATC CTGGTATTGC AACAATACTC GGCTTTTTGC ACTGACGGAG AGAAGAGTCA 300
TTCTTGACAC CATTGCCTGC AATTTTACTT GTGTGGAACA ATCTGGGCAT CGACAGAGCA 360
TTTGGATTAC ATGGCGTGCA CAACCTGTCT TACAAACCTT GTGTGCACAG CCATCAAAAC 420
CAGTCACTTG TGGTCAGCAT GTTACTTTGT ATTGTTCTAC CTCTGGAAAT AATGTTACCG 480
TTTGGCATCT ACCAAACGGA CGAAATGAAA CCGTGTGACA AACTAAATAC TATAATTTTA 540
CGCTGATGAG CCAAACTGAG GGGTGTATA CTGTTTCTAA CCGGCTGTCT TCTCGCCTGT 600
CAAATCGTAT ATGTTTTTGG GCGCGTTGTG CCAATATAAC TCCAGAACT CATACTGTAT 660
CTGTCAGCAG TACTACAGGC TTTAGAACAT TGAGTACTAA TAGCTTAGTG AAGATAATCC 720
ATGCAACCCAC ACGTGATGTA GTTGTAGTGA AAGAAGCAAA ATCTACACAT TTTGATATTG 780
AAGTGCATTT TTTTGTATTT ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT 840

```

|             |            |             |            |             |            |      |
|-------------|------------|-------------|------------|-------------|------------|------|
| TAGGAAGTAT  | TATGTTTGGC | CATTGTGAAA  | AACAAGCTGA | CTCAACAAA   | ACAGTGGCAG | 900  |
| AACAATTGCA  | GGATTATTAT | TCCCTACAGG  | ATTTGTGCAC | GGAAAGACTAT | ACGCAACCAG | 960  |
| TGGATTGGTA  | CTGACATTCA | GGTAAGATAA  | TCTAAATATT | CTCTATAACA  | TAATTGTAAT | 1020 |
| GTGTTTTATG  | TTTATAGCTA | CAAAATGTTT  | ATGCAAAATA | CATTTTATGA  | GGTGGGATAC | 1080 |
| TTATTAAAG   | CATTGTCTTA | AGTACATTAA  | AAGGACATTG | TATAACCGTG  | CTACTTACAG | 1140 |
| CATGGCCTTT  | TTAAGACAAA | CACTGTGGAT  | TTTATGGACA | TTTACCATGG  | TTATTGGCCA | 1200 |
| GGACAATGAA  | AAGTGTTCGC | AAAAAACCTT  | AATTGGATAT | AGACTTAAAA  | TGTCTCGTGA | 1260 |
| CGGTGACATT  | GCAGTTGAG  | AAACAGTGGG  | ATTACGTTGT | AGATCTGGAT  | ACACTACTTA | 1320 |
| TGCCCCGAAT  | ATAACAGCAA | CATGTTTACA  | AGGTGGGACG | TGGTCTGAAC  | CAACGGCAAC | 1380 |
| ATGTAACAAA  | AAGTCTGTGC | CAAACCCAGG  | TGAAATACAA | AATGGAAAGG  | TTATATTTCA | 1440 |
| TGSTGGACAA  | GATGCCTTAA | AATATGGGGC  | AAACATTTCA | TATGTTTGTA  | ATGAAGGATA | 1500 |
| TTTTTTGGTT  | GGTGGAGAA  | AGGTGGGATA  | TTGTATGATT | GGAGCATCTG  | GCCAAATGGC | 1560 |
| GTGGTCATCT  | TCTCTCTCTT | TTTGTGAAAA  | AGAAAAGTGT | CACAGACCGA  | AAATCAAAAA | 1620 |
| TGGAGATTTT  | AAGCCTGATA | AAGATTATTA  | TGAGTATAAT | GATGCAGTTC  | ATTTTGAATG | 1680 |
| TAATGAAGGA  | TATACTCTAG | TTGGACCACA  | TTCCATTGCA | TGTGCAGTTA  | ATAACAGCTG | 1740 |
| GACATCTAAC  | ATGCCAACCT | GTGAACTCGC  | AGGCTGTAAA | TTTCCATCGG  | TGACTCATGG | 1800 |
| TTATCCAAATC | CAAGSTTTTT | CTCTTACTTA  | TAAACATAAG | CAAAGTGTTA  | CTTTTGCATG | 1860 |
| CAATGATGGA  | TTTGTCTCTA | GAGGATCCCC  | CACAATTACG | TGTAACGTTA  | CTGAATGGGA | 1920 |
| CCCACCACTT  | CCTAAGTGTG | TTTTGGAAGA  | TATAGATGAT | CCAAACAAAT  | CAATCCTTGG | 1980 |
| ACGTTTGCAT  | CCAACACCCA | ATGAAAAACC  | AAATGGTAAT | GTCTTTCAAC  | GCTCAAACTA | 2040 |
| TACAGAACCT  | CCAACAAAGC | CTGAAGACAC  | CCATACAGCA | GCTACTTGTG  | ATACCAACTG | 2100 |
| TGAACAGCCA  | CCTAAATTC  | TGCCAACATC  | CGAAGGTTTT | AATGAGACTA  | CCACATCTAA | 2160 |
| TACAATTACA  | AAACAATTAG | AGGATGAGAA  | AACTATATCC | CAGCCAAATA  | CACATATTAC | 2220 |
| ATCTGCCTTA  | ACATCCATGA | AAGCGAAAGG  | TAACTTTACC | AACAAGACCA  | ATAACTCTAC | 2280 |
| TGATCTACAT  | ATAGCGTCTA | CACCCACTTC  | CCAAGATGAT | GCTACGCCCT  | CAATACCTAG | 2340 |
| TGTACAGACA  | CCCAATTATA | ATACTAACGC  | ACCGACACGT | ACACTAACGT  | CTCTCCATAT | 2400 |
| TGAAGAAGGC  | CCATCCAATT | CTACTACTTC  | AGAAAAGGCC | ACTTCTCTTA  | CTCTCTCACA | 2460 |
| CAACTCACAC  | AAAAATGACA | CCGGAGGCAT  | ATACACAACA | TTAAACAAAA  | CAACACAGTT | 2520 |
| GCCATCCACT  | AATAAACCTA | CAACACGTCA  | AGCCAAGAGT | TCCACTAAGC  | CACGCGTTGA | 2580 |
| GACACACAAT  | AAAACAACCA | GTAATCCTGC  | CATTTCTTTA | ACAGATTCTG  | CAGATGTGCC | 2640 |
| TCAGAGACCG  | CGAGAACCAA | CACTCCCTCC  | CATTTTCAGG | CCACCGGCGT  | CTAAAAATCG | 2700 |
| CTATCTGGAA  | AAGCAACTAG | TTATTGGACT  | ACTAACCSCT | GTGCGCCTAA  | CGTGTGGACT | 2760 |
| GATTACCTTA  | TTTCACTATC | TGTTCTTTTCG | TTAGCCTAGA | ACTTGCTCCA  | GTGTTAGACA | 2820 |
| GGGCTATGAT  | TGCTTCTCCA | CGCTGTCCAC  | CTTAACACTT | CCCAATAACA  | AATCCGGTAT | 2880 |

|            |            |            |             |            |            |      |
|------------|------------|------------|-------------|------------|------------|------|
| GCAGCAGCGT | GACACTACTA | ATGTAACCTA | AAAAATGTGC  | ATGTSSTATG | TATTGTACTA | 3340 |
| AAGATACCGA | CCAATACAAG | ACAACTAATA | TTAACCATAG  | TGTGCGTTTC | TTTGTATAAA | 3350 |
| ATACGCGTGT | GGGAAAGCGA | CAGAAGGGGG | CGGCGTTTCC  | ATATGAGGCC | AAGTGCAATG | 3360 |
| GCTATTTTAA | GGGCGGTGAC | CACGCACTAT | AGTGCGCGGT  | GTGSCAGAAA | ATTCAACAGG | 3370 |
| TATATAAACA | AGGAAAGGGG | ACTCTGCGCG | CTTAAGCGCC  | AAGCCATTAT | ACACACGGGT | 3380 |
| TTTTTGTTGT | CTTGCGCAAT | CGTGTCTCCA | TGGCGCTAAA  | GGGACCACAA | ACCGTCGAGG | 3390 |
| AAAATATTGG | GTCTGCGGGC | CCCCTGGGTC | CCTGCGGGTA  | CCTCTATGCC | TATCTGACAC | 3400 |
| ACAACTTCCC | CATAGGGGAA | GCCTCCCTGC | TGGGCAATGG  | CTACCCGGAG | GCAAAATAT  | 3410 |
| TTTCACTACC | TCTTTTGCAC | GGGCTCACAG | TGGAATCCGA  | TTTCCCTTAA | AACGTAAAGG | 3420 |
| CGGTGCACAA | GAAATCGAT  | GCAACCACAG | CTTCTGTGAA  | ATTAACTTCA | TACCACAGGG | 3430 |
| AGGCCATCGT | CTTTCATAAT | ACTCACTTAT | TTCAGCCAAT  | CTTTCAGGGA | AAGGGACTGG | 3440 |
| AAAAGTTATG | TGAGAGAGAG | CGAGAGCTGT | TTGGATTTC   | AACGTTTGT  | GAGCAACAAC | 3450 |
| ACAAAGGGAC | GCTCTGGAGC | CCAGAGGCAT | GCCTCAGCT   | ACCGTGCGCG | AATGAGATTT | 3460 |
| TTATGGCGGT | CATAGTTACA | GAGGGATTCA | AGGAGAGACT  | GTACGGCGGG | AAACTGCTGC | 3470 |
| CCGTGCCCTC | TCAGACAACG | CCCCTACACA | TTGGGGAAACA | CCAGGCGTTC | AAGATACCTT | 3480 |
| TGTATGACGA | GGATCTGTTT | GGTCCAAGTC | GCGCCCAAGA  | ACTATGTAGG | TTTACCAACC | 3490 |
| CCGATATCAG | TAGATACCTA | CATGACTCCA | TATTCACTGG  | AATAGCACAG | GCTCTAAGGG | 3500 |
| TAAAGGACGT | TAGCACGGTC | ATCCAAGCCT | CAGAAAGGCA  | ATTTGTGCAC | GACCAATACA | 3510 |
| AGATACCAAA | GCTGGTCCAA | GGCAAGGACT | TCCCCAGTG   | TGCTTCCAGG | GGAACTGACG | 3520 |
| GCTCTACCTT | AATGGTGATA | GACASTCTGG | TGGCTGAACT  | TGGTATGAGT | TATGCTCTGT | 3530 |
| CGTTTATTGA | GGGACCCCGG | GATAGCTGGG | AGGTTCTAAA  | TTATGACACG | TGGCCCATCT | 3540 |
| TTGAAAAGTG | CGAGACGCCA | GATGCCCGCC | TTCTGTCACT  | AGAAGTTTGG | CACGCAGAGG | 3550 |
| AGGCCTTGCA | TATTGGCGCC | CAGCTGTTTG | CGGCCAACTC  | TGTGCTCTAC | CTGACCAGAG | 3560 |
| TGGCAAGGCT | GCCTCAGAAG | AATCAGAGAG | GAGACGCCAA  | CATGTACAAC | TCATTCTACC | 3570 |
| TACAGCATGG | CCTGGGATAC | CTCTCAGAGG | CAACAGTAAA  | GGAAAATGGA | GCCTCTGGCT | 3580 |
| TCAAGGGCGT | GCCAGTGTCT | GCACTGGATG | GGTCATCTTA  | CACCTCCAG  | CACCTGGCCT | 3590 |
| ACGCTCTCTC | TTTCTCCCCA | CATCTCCTGG | CAAGGATGTG  | TTACTATCTG | CAGTTCTTGG | 3600 |
| CCCACCATAA | AAACACCAAC | AGTCAGTCAT | ACAATGTGGT  | GGACTACGTG | GGCAGCGCGG | 3610 |
| CACCTAGTCA | AATGTGTGAC | CTGTGTGAGG | GGCAATGTCC  | AGTTGTATGC | ATCAACAGGC | 3620 |
| TGTTTTACAG | GATGAAGGAC | AGTTCCCGAC | CTGTTCTGTG  | AAACGTTAAG | AGAGACCCAT | 3630 |
| ATGTGATCAC | GGGCACAGCG | GGAACGTACA | ATGACCTAGA  | GATTCTCGGA | AACTTTGCCA | 3640 |
| CCTTCAGGGA | GAGAGAGGAG | GAGGGGAATC | CTGTGGAAGA  | TGCTCCAAAG | TATACATATT | 3650 |
| GGCAACTATG | CCAGAATATA | ACCGAGAAGC | TAGCCTCCAT  | GGGCATCTCG | GAGGGCGGCG | 3660 |
| ATGCCCTAAG | AACCTTCATT | GTGGACATCC | CCAGCTTCCT  | CAAAGTGTTC | AAGGGGATAG | 3670 |

|             |             |            |            |             |            |      |
|-------------|-------------|------------|------------|-------------|------------|------|
| ACAGCAGCGT  | AGAGGCAGAG  | CTCCTAAAGT | TTATTAACTG | CATGATCAAA  | AACAATTACA | 4980 |
| ACTTCAGAGA  | GAACATCAAA  | TCGCTCCATC | ACATCCTTCA | GTTTGCATGC  | AACGTATACT | 5040 |
| GGCAGGCGCC  | GTGCCCCGTT  | TTTCTGACCC | TTTACTACAA | GTCACTGCTG  | ACGGTCATAC | 5100 |
| AGGACATATG  | TCTGACGTCA  | TGTATGATGT | ACGAGCAGGA | CAACCCGGCC  | GTGGGAATTC | 5160 |
| TACCATCCGA  | GTGGCTTAAA  | ATGCACCTTC | AGACAATGTG | GACCAACTTC  | AAGGGTGCCT | 5220 |
| GCTTCGACAA  | AGGAGCAATC  | ACGGGCGGGG | AACTAAAAAT | AGTCCACCAG  | TCCATGTTCT | 5280 |
| GTGACCTCTT  | TGACACCGAC  | GCTGCCATAG | GAGGGATGTT | TGCACCCGCT  | CGGATGCAGG | 5340 |
| TCAGGATAGC  | CAGAGCAATG  | CTCATGGTTC | CAAAAACCAT | AAAAATAAAA  | AACAGGATCA | 5400 |
| TCTTTTCCAA  | CTCCACCGGA  | GCAGAGTCGA | TCCAGGCAGG | TTTTATGAAG  | CCGCCCAGCC | 5460 |
| AAAGGGATTG  | ATACATCGTC  | GGAGGACCCT | ACATGAAATT | CCTAAACGCC  | CTGCACAAAA | 5520 |
| CACTTTTTCC  | TTCCACAAAA  | ACTTCTGCCC | TGTACTTGTG | GCATAAGATT  | GGCCAGACCA | 5580 |
| CAAAAAATCC  | CATACTACCA  | GGTGTCTCGG | GGGAACACCT | AACGGAGTTA  | TGTAATTATG | 5640 |
| TAAAGGCAAG  | TAGCCAGGCT  | TTGGAAGAGA | TAAATGTTTT | GGACCTTGTG  | CCAGACACCC | 5700 |
| TGACATCATA  | TGCGAAAAATA | AAACTAAACA | GTTCCATTCT | CCGGGCTTGC  | GGACAGACAC | 5760 |
| AGTTTTATGC  | AACTACTCTC  | TCTTGCCCTT | CGCCAGTGAC | TCAGCTGCTT  | CCGCCCAGGG | 5820 |
| AGTACCCCCA  | CGTACTGGGG  | CCAGTGGGGT | TGTCATCTCC | AGATGAATAC  | AGGGCAAAAG | 5880 |
| TGCGCGSCAG  | GTCTGTAACC  | ATTGTACAGT | CAACACTGAA | GCAAGCTGTT  | TCCACCAACG | 5940 |
| GACGACTCCG  | GCCTATCATT  | ACCCTGCCAC | TGGTGGTCAA | CAAAATATACA | GGGAGCAACG | 6000 |
| GGAAACACAA  | CGTCTTTCAC  | TGTGCAAAAC | TGGGATACTT | CTCGGGGAGA  | GGGGTGGACA | 6060 |
| GAAATCTCAG  | GCCAGAAAGC  | GTCCCTTTTA | AAAAGAATAA | TGTCAGCTCT  | ATGCTAAGAA | 6120 |
| AACCCACCT   | GATTATGACC  | CCCCTGGTAG | ACAGGCTGCT | AAAGAGAATA  | GTTGGCATCA | 6180 |
| ACTCTT333GA | ATTGAGGGCA  | GAAGCGGTTA | AGAGAAGTGT | GCAGAATGTC  | CTGGAAGACA | 6240 |
| GAGATAACCC  | AAACCTGCCG  | AAGACAGTTG | TATTAGAGTT | GSTTAAGCCA  | CCTCGGTGGA | 6300 |
| GCTCCTGTGC  | AAGTCTCACA  | GAGGAGGACG | TGATTTACTA | CCTGGGCCCT  | TATGCCGTAC | 6360 |
| TTGGGGACGA  | GCTCCTGTCA  | TTACTGAGCA | CAGTGGGCCA | GGCGGGGGTG  | CCATGGACGG | 6420 |
| CCGAGGGTGT  | GGCCTCGGTC  | ATCCAGGACA | TAATAGATGA | TTGCGAGTTA  | CAGTTTGTGG | 6480 |
| GCCCAGAAGA  | GCCTTGCCCT  | ATCCAAGGAC | AGTCGGTAGT | GGAGGAGCTT  | TTTCCGTCCC | 6540 |
| CGGGCTTCCC  | AAGCCTGACA  | GTGGGTAAAA | AACGAAAAAT | CGCATCCCTG  | CTCTCTGACC | 6600 |
| TGGATTTGTA  | GTTGTGTACC  | CGTAACGATG | GCAAAGGAAC | TGGCGGCGGT  | CTATGCCGAT | 6660 |
| GTSTCAGCCC  | TAGCCATGGA  | CCTCTGTCTT | CTTAGTTACG | CAGACCCGGC  | AACACTGGAC | 6720 |
| ACTAAAAGTC  | TGGCCCTCAC  | TACAGGGAAG | TTTCAGAGCC | TTACCGGCAC  | ACTACTCCCC | 6780 |
| CTCCTCAGAC  | GACAAAACGC  | ACACGAATGC | TCAGGTCTGT | CACTAGAATT  | GGAGCACTTT | 6840 |
| TGGAAAACGT  | GGCTGATGCT  | CTGGCCACGT | TGGGAGTGTG | CACTAGCAGA  | AAACTGTCTC | 6900 |
| CAGAAGAGCA  | TTTTTCCCTC  | CTGCATTTGG | ACACAACATG | CAACAAGCAA  | CCGGAGCGTT | 6960 |

|             |            |            |            |            |            |      |
|-------------|------------|------------|------------|------------|------------|------|
| AGGTTTAATT  | TTTACGGAAA | TTGGGCGTTG | GAGTTAAAGC | TGTCACATAA | AAACGACGTT | 7010 |
| GAAATTTTCT  | TTAAACGTCT | TAGTAGCGTT | TTTTATTGTA | TAGGATCGGG | CAGTGCCTCT | 7080 |
| GAGGTTTTAG  | GGGAGGTATT | GCCTTTCGTT | GGGAAGCTGA | GGGGTATCTC | ACCCGTACCT | 7140 |
| GGGCGGAGCT  | TATATGTCTC | AAATCTGCCC | TGCCTAGAAT | GCCTTCAGGA | AGTGTGTCTC | 7200 |
| ACTCCCAACC  | AGGGCACCAG | TCTGCAGGCG | ATGCTCCGAG | ACACGGCGCT | CAGTCACATA | 7260 |
| TGTACCCCCG  | CATGCGGTGA | GCCTGTCCGG | GGCCTCTTTG | AGAACGAGCT | AAAACAGCTC | 7320 |
| GGGCTTCAAA  | CCCTTGAGTC | CATACCTACT | ACCCCTGTCT | AGTCCCGGGT | AAGGCAAGAT | 7380 |
| GATGAAATCA  | GACAGAGCTC | TCTAATGGCG | GTAGGAGATC | ACCACATTTT | CGGAGAAGTG | 7440 |
| ACCAGATCTG  | TCTTGGAAT  | CTCAACCTG  | ATCTATTGGA | GCTCTGCCCC | CTCGGATGCC | 7500 |
| ACCTGCGAGG  | GAGACAGAGA | CTGCTCTCAC | CTGGCCTCGC | TGTTTACTCA | CGAGGCTGAC | 7560 |
| ATGCATAAAA  | GGCGCGTCGA | CCTGGCGGGA | TGCTTGCGCG | AACGCGGCAC | GCCCCAACAC | 7620 |
| TTTTTTGACT  | GCTTTCGCCC | AGACTCCCTA | GAAACCCCTT | TCTGTGGTGG | TCTTTTTAGC | 7680 |
| TCCGTGGAGG  | ACACCATAGA | AAGTCTCCAA | AAGGACTGCT | CTTCTGCCCT | CTACCAACAG | 7740 |
| GTAAACTACA  | CTACTGCACT | GCAAAAACAG | AACGAGTTTT | ACGTCCGACT | CAGCAACTG  | 7800 |
| CTGGCAGCTG  | GTCAGCTAAA | TTTGGGCAAA | TGTTCCACTG | AAAGTTGCCA | ATCCGAGGCC | 7860 |
| CCTAGGCAGC  | TGGTAGGTGG | GGGCAAAACC | GAGGAAGTGC | TGAGGGATGC | AAAACACCGG | 7920 |
| CAAGAACTAT  | ACCTTCAGAA | AGTGGGCACG | GACGGTTTTA | AAAAACTCTC | TGATTGTATA | 7980 |
| AGACACCAGG  | GCCACATCCT | GTCTCAGACC | CTGGGTCTAA | GACTGTGGGG | GTCTGTCTAT | 8040 |
| TACAACGAGG  | CATCTGCCCT | ACAAAACGAC | TTTTTACACA | GAGCAGAGTT | CATATCCCTC | 8100 |
| CCCTGGCAGG  | ACCTGACGGT | CGACTGTCCA | ACGCGGTTTG | AAAATTCTAA | ATATATCAAA | 8160 |
| AATTCTCTGT  | ACTGCCAGCG | TCTGGGCGCG | GAACACGTAG | AGATCCTGAC | ACTGGAGTTC | 8220 |
| TACAACTTTA  | TCACGGGCCC | GCTGTCAAAG | CGACATACTT | TATTTCCGAG | TCTCCCAAT  | 8280 |
| GTGACGCTGG  | CTCAGTGCTT | CGAGGCTGCG | GGCATGCTTC | CCCATCAAAA | GATGATGGTA | 8340 |
| TCAGAGATGA  | TCTGGCCGAG | CATAGAGCCG | AAGGACTGGA | TAGAGCCCCA | CTTCAACCAG | 8400 |
| TTCTATAGCT  | TTGAGAATCA | AGACATAAAC | CATCTGCAAA | AGAGAGCTTG | GGAAATATAT | 8460 |
| AGAGAGCTGG  | TATTATCGGT | TTCTCTGTAC | AACAGAACTT | GGGAGAGGGA | GCTAAAAATA | 8520 |
| CTTCTCACGC  | CTCAGGGCTC | ACCGGGGTTT | GAGGAACCGA | AACCCGCGAG | ACTCACAAAG | 8580 |
| GGGCTGTACC  | TAACATTTGA | GACATCTGCG | CCCTTGGTGT | TGGTGGATAA | AAAATATGGC | 8640 |
| TGGATATTTA  | AAGACCTGTA | CGCCCTTCTG | TACCACCACC | TGCAACTGAG | CAACCACAA  | 8700 |
| GACTCCGAGG  | TCTAGATTGG | CCACCCCTGG | GACTGTCTAT | CTGTTGGTCT | GTTTTTGGCG | 8760 |
| AGGCGCGGCG  | CACTCGAGGG | GTGACACCTT | TCAGACGTTC | AGTTCCCCCA | CACCCCGAGG | 8820 |
| ATCTTCCCTCT | AAGGCCCCCA | CCAAACCTGG | TGAGGAAGCA | TCTGGTCCCT | AGAGTGTGGA | 8880 |
| CTTTTACCAG  | TTCAAGAGTG | GATGTGCATC | GATCACCGGG | GAGCTTTTTT | GSTTCAACCT | 8940 |
| GGAGCAGACG  | TGCCCGAGAC | CCAAAGACAA | GTACCACCAA | GAAGGAATTT | TACTGCTGTA | 9000 |



|            |             |            |             |             |             |       |
|------------|-------------|------------|-------------|-------------|-------------|-------|
| CAAAAAAAG  | ATAGTGCCTC  | ATATCTTTAA | GSTGCGGCGC  | TATAGGAAAA  | TTGCCACCTC  | 9060  |
| TGTCACGGTC | TACAGGGGCT  | TGACAGAGTC | CGCCATCACC  | AACAAGTATG  | AACCTCCGAG  | 9120  |
| ACCCGTGCCA | CTCTATGAGA  | TAAGCCACAT | GGACAGCACC  | TATCAGTGCT  | TTAGTTCCAT  | 9180  |
| GAAGGTAAAT | GTCACGGGG   | TAGAAAACAC | ATTTACTGAC  | AGAGACGATG  | TTAACACCCAC | 9240  |
| AGTATTCTCT | CAACCAAGTAG | AGGGGCTTAC | GGATAACATT  | CAAAGGTACT  | TTAGCCAGGC  | 9300  |
| GSTCATCTAC | GCGGAACCCG  | GCTGGTTTTG | CGGCATATAC  | AGAGTTAGGA  | CCACTGTCAA  | 9360  |
| TTGCGAGATA | GTGGACATGA  | TAGCCAGGTC | TGCTGAACCA  | TACAATTACT  | TTGTCAAGTC  | 9420  |
| ACTGGGTGAC | ACGGTGGAG   | TCTCCCTTTT | TTGCTATAAC  | GAATCCTCAT  | GCAGCACAAAC | 9480  |
| CCCCAGCAAC | AAAAATGGCC  | TTAGCGTCCA | AGTAGTTCTC  | AACCACACTG  | TGGTCAAGTA  | 9540  |
| CTCTGACAGA | GGAAACAGTC  | CCACTCCCCA | AAACAGGATC  | TTTGTGGAAA  | CGGGAGCGTA  | 9600  |
| CACGCTTTTG | TGGGCTTCCG  | AGAGCAAGAC | CACGGCCGTG  | TGTCCGCTGG  | CACTGTGGAA  | 9660  |
| AACCTTCCCG | CGCTCCATCC  | AGACTACCCA | CGAGGACAGC  | TTCCACTTTG  | TGGCCAAACGA | 9720  |
| GATCACGGCC | ACCTTCACGG  | CTCTCTAAC  | GCCAGTGGCC  | AACCTTTACG  | ACACGTACTC  | 9780  |
| TTGTCTGACC | TGGATATCA   | ACACCACGCT | AAACGCCAGC  | AAGGCCAAAC  | TGGCGAGCAC  | 9840  |
| TCACGTCCCT | AACGGGACGG  | TCCAGTACTT | CCACACAACA  | GGCGGACTCT  | ATTTGGTCTG  | 9900  |
| GCAGCCCATG | TCCCGCATTA  | ACCTGACTCA | CGCTCAGGGC  | GACAGCGGGA  | ACCCCAAGTC  | 9960  |
| ATCGCCGCCC | CCCTCCGCAT  | CCCCCATGAC | CACCTCTGCC  | AGCCGCAGAA  | AGAGACGGTC  | 10020 |
| AGCCAGTACC | GCTGCTGCCG  | GCGGCGGGGG | GTCCACGGAC  | AACCTGTCTT  | ACACGCAGCT  | 10080 |
| GCAGTTTGGC | TACGACAAAC  | TGCGGGATGG | CATTAAATCAG | GTGTTAGAAG  | AACCTCTCCAG | 10140 |
| GGCATGGTGT | CGCGAGCAGG  | TCAGGGACAA | CCTAATGTGG  | TACGAGCTCA  | GTAAATCAA   | 10200 |
| CCCCACCAGC | GTTATGACAG  | CCATCTACGG | TGCACCTGTA  | TCCGCCAAGT  | TGCTAGGAGA  | 10260 |
| CGCCATTTCC | GTGACCGAGT  | GCATTAAAGT | GGACCAGAGC  | TCCGTAAACA  | TCCACAAGAG  | 10320 |
| CCTCAGAAAC | AATASTAAGG  | ACGTGTGTTA | CGCGCGCCCC  | CTGGTGACGT  | TTAAGTTTTT  | 10380 |
| GAACAGTTCC | AACCTATTCA  | CCGGCCAGCT | GGGCGCGCGC  | AATGAGATAA  | TACTGACCAA  | 10440 |
| CAACCAGGTG | GAAACCTGCA  | AAGACACCTG | CGAACACTAC  | TTCAATCACC  | GCAACGAGAC  | 10500 |
| TCTGGTGTAT | AAGGACTACG  | CGTACCTGCG | CACTATAAAC  | ACCACTGACA  | TATCCACCTT  | 10560 |
| GAACACTTTT | ATCGCCCTGA  | ATCTATCCTT | TATTCAAAC   | ATAGACTTCA  | AGGCCATCGA  | 10620 |
| GCTGTACAGC | AGTGCAGAGA  | AACGACTCGC | GAGTAGCGTG  | TTTGACCTGG  | AGACGATGTT  | 10680 |
| CAGGGAGTAC | AACCTACTACA | CACATCGTCT | CGCGGGTTTG  | CGCGAGGATC  | TGGACAACAC  | 10740 |
| CATAGATATG | AACAAGGAGC  | GCTTCGTAAG | GGACTTGTGG  | GAGATAGTGG  | CGGACCTGGG  | 10800 |
| TGGCATCGGA | AAACCGGTGG  | TGAACGTGGC | CAGCAGCGTG  | GTCACCTCTAT | GTGGCTCATT  | 10860 |
| GSTTACCGGA | TTCAATAATT  | TTATTAAACA | CCCCCTAGGT  | GGCATGGTGA  | TGATCATTAT  | 10920 |
| CGTTATAGCA | ATCATCCTGA  | TCATTTTTAT | GCTCAGTCGC  | CGCACCATAA  | CCATAGCCCCA | 10980 |
| GGCGCCGGTG | AAGATGATCT  | ACCCCGACGT | AGATCGCAGG  | GCACCTCTTA  | GCGGCGGAGC  | 11040 |

|             |             |            |            |            |             |       |
|-------------|-------------|------------|------------|------------|-------------|-------|
| CCCCAACACGG | GAGGAATCA   | AAACATCCT  | GCTGGGAATG | CACCAGCTAC | AACAGAGAG   | 11150 |
| GAGGCAGAAG  | GCGGATGATC  | TGAAAAAAG  | TACACCTCG  | GTGTTTCAGC | GTACCGCAAA  | 11160 |
| CGGCCTTGGT  | CAGCGTCTGA  | GAGGATATAA | ACCTCTGACT | CAATCGCTAG | ACATCAGTCC  | 11170 |
| GGAAACGGGG  | GAGTGACAGT  | GGATTCGAGG | TTATTGTTTG | ATGTAAATTT | AGGAACACCG  | 11180 |
| GCCCCCTCTCT | GAAGCACCCAC | ATACAGACTG | CAGTTATCAA | CCCTACTCGT | TGCACACAGA  | 11190 |
| CACAAATTAC  | CGTCCGCAGA  | TCATGGATTT | TTTCAATCCA | TTTATCGACC | CAACTCGCGG  | 11200 |
| AGGCCCCGAGA | AACACTGTGA  | GGCAACCCAC | GCCGTCACAG | TCGCCAAGTG | TCCCCCTCGA  | 11210 |
| GACAAGAGTA  | TGCAGGCTTA  | TACCGGCTG  | TTTCCAAACC | CCGGGGCGAC | CCGGCGTGST  | 11220 |
| TGCGCTGGAC  | ACCACATTTG  | CACCCACCTA | CTTCAGGGG  | CCCAAGCGGG | GAGAAATATT  | 11230 |
| CGCGGGAGAG  | ACTGGGTCTA  | TCTGGAAAC  | AAGGCGCGGA | CAGGCACGCA | ATGCTCTAT   | 11240 |
| GTCCGACCTC  | ATATTCCAGG  | TATACGACT  | CGTGGAGACC | ACCTACACGG | CCGACCGCTG  | 11250 |
| CGAGGACGTG  | CCATTTAGCT  | TCCAGACTGA | TATCATTCCT | AGCGGCACCG | TCTCAAGCT   | 11260 |
| GCTCGSCAGA  | ACACTAGATG  | GGGCCAGTGT | CTGCGTGAAC | GTTTTTCAGG | AGCGCTGCTA  | 11270 |
| CTTCTACACA  | CTAGCACCTC  | AGGGGGTAA  | CCTGACCCAC | GTCTCCAGC  | AGGCCCTCCA  | 11280 |
| GGCTGGCTTC  | GGTGGCGCAT  | CCTGCGGCTT | CTCCACCGAG | CCGGTCAGAA | AAAAATCTTT  | 11290 |
| GCGCGCGTAC  | GACACACAA   | AATATGCTGT | GCAAAAAATA | ACCCTGTCT  | CCAGTCCGAT  | 12000 |
| GATCGCAAGC  | CTTAGCGAGC  | GCCTAACAA  | CTGTGGGTGC | GAGGTGTTTG | AGTCCATGT   | 12010 |
| GGACGCCATT  | AGGCGCTTCG  | TGCTGGACCA | CGGGTTCTCG | ACATTGGGGT | GGTACGAGTG  | 12020 |
| CAGCAATCCG  | GCCCCCGCGA  | CCCAGGCCAG | AGACTCTTGG | ACGGAACTGG | AGTTTCACTG  | 12030 |
| CAGCTGGGAG  | GACCTAAAGT  | TTATCCCGGA | GAGGACGGAG | TGGCCCCCAT | ACTCAATCCT  | 12040 |
| ATCCTTTGAT  | ATAGAAATGA  | TGGGCGAGAA | GGGTTTTCCT | AACGCGACTC | AAGACGAGGA  | 12050 |
| CATGATTATA  | CAAACTCTCT  | GTGTTTTACA | CACAGTCGGC | AACGATAAAC | CGTACACCCG  | 12060 |
| CATGCTACTG  | GGCCTGGGGA  | CATGCGACCC | CCTTCTTGGG | GTGGAGGTCT | TTGAGTTTCC  | 12070 |
| TTCCGAGTAC  | GACATGCTGG  | CCGCCTTCTT | CAGCATGCTC | CGCGATTACA | ATGTGGAGTT  | 12080 |
| TATAACGGGG  | TACAACATAG  | CAAACTTTGA | CCTTCCATAC | ATCATAGCCC | GGGCAACTCA  | 12090 |
| GGTGTACGAC  | TTCAAGCTGC  | AGGACTTCAC | CAAAATAAAA | ACTGGGTCCG | TGTTTGAGGT  | 12100 |
| CCACCAACCC  | AGAGGCGGTT  | CCGATGGGGG | CAACTTCATG | AGGTCCCACT | CAAAAGGTCAA | 12110 |
| AATATCGGGG  | ATCGTCCCCA  | TAGACATGTA | CCAGGTTTGC | AGGGAAAAGC | TGAGTCTGTC  | 12120 |
| AGACTACAAG  | CTGGACACAG  | TGGCTAAGCA | ATGCCTCGGT | CGACAAAAAG | ATGACATCTC  | 12130 |
| ATACAAGGAC  | ATACCCCGGC  | TTTTTAAATC | TGGGCTTGAT | GGTGGCGCAA | AGGTGGGAAA  | 12140 |
| CTACTGTGTT  | ATTGACTCGG  | TCCTGGTTAT | GGATCTTCTG | CTACGGTTTC | AGACCCATGT  | 12150 |
| TGAGATCTCG  | GAAATAGCCA  | AGCTGGCCAA | GATCCCCACC | CGTAGGGTAC | TGACGGACGG  | 12160 |
| CCAACAGATC  | AGGGTATTTT  | CGTGCCTCTT | GGAGGCTGCT | GCCACGGAAG | GTTACATTCT  | 12170 |
| CCCCGTCCCA  | AAAGGAGACG  | CGGTTAGCGG | GTATCAGGGG | GCCACTGTAA | TAAGCCCTTC  | 12180 |

|             |            |            |            |            |             |       |
|-------------|------------|------------|------------|------------|-------------|-------|
| TCCGGGATTTC | TATGACGACC | CCGTACTCGT | GGTGGATTTT | GCCAGCTTGT | ACCCCAAGTAT | 13140 |
| CATCCAAAGCG | CACAACTTGT | GCTACTCCAC | ACTGATACCC | GGCGATTCCG | TCCACCTGCA  | 13200 |
| CCCACACCTC  | TCCCCGGACG | ACTACGAAAC | CTTTGTCTTC | AGCGGAGGTC | CGGTCCACTT  | 13260 |
| TGTAAAAAAA  | CACAAAAGGG | AGTCCCTTCT | TGCCAAGCTT | CTGACGGTAT | GGCTCGCGAA  | 13320 |
| GAGAAAAGAA  | ATAAGAAAGA | CCCTGGCCTC | ATGCACGGAC | CCCGCACTGA | AAACTATTCT  | 13380 |
| AGACAAACAA  | CAACTGGCCA | TCAAGGTTAC | CTGCAACGCC | GTTTACGGCT | TCACGGGCGT  | 13440 |
| TGCTCTGGC   | ATACTGCCCT | GCCTAAACAT | AGCGGAGACC | GTGACACTAC | AAGGGCGAAA  | 13500 |
| GATGCTGGAG  | AGATCTCAGG | CCTTTGTAGA | GGCCATCTCG | CCGGAACGCC | TAGCGGCTCT  | 13560 |
| CCTGCGGAGG  | CCAATAGACG | TCTCACCCTA | CGCCCGATTG | AAGGTCTAT  | ACGGCGACAC  | 13620 |
| TGACTCTCTT  | TTCATATGCT | GCATGGGTTT | CAACATGGAC | AGCCTGTCAG | ACTTCGCGGA  | 13680 |
| GGAGCTAGCG  | TCAATCACCA | CCAACACGCT | GTTTCGTAGC | CCCATCAAGC | TGGAGGCTGA  | 13740 |
| AAAGATCTTC  | AAGTGCCTTC | TGCTCCTGAC | TAAAAAGAGA | TACGTGGGGG | TACTCAGTGA  | 13800 |
| CGACAAGGTT  | CTGATGAAGG | GCGTAGACCT | CATTAGGAAA | ACAGCCTGTC | GTTTTGTCCA  | 13860 |
| GGAAAAGAGC  | AGTCAGGTCC | TGGACCTCAT | ACTGCGGGAG | CCGAGCGTCA | AGGCCGCGGC  | 13920 |
| CAAGCTTATT  | TGGGGGCAGG | CGACAGACTG | GCTGTACAGG | GAAGGGCTCC | CAGAGGGGTT  | 13980 |
| CCTCAAGATA  | ATTCAAGTGC | TCAACGCGAG | CCACCGGGAA | CTGTGCGAAC | GCAGCGTACC  | 14040 |
| AGTAGACAAA  | CTGACGTTTA | CCACCGAGCT | AAGCCGCCCG | CTGGCGGACT | ACAAGACGCA  | 14100 |
| AAACCTCCCG  | CACCTGACCG | TGTACCAAAA | GCTACAAGCT | AGACAGGAGG | AGCTTCCACA  | 14160 |
| GATACACGAC  | AGAATCCCTT | ACGTGTTCTG | CGACGCCCCA | GGTAGCCTGC | GCTCCGAGCT  | 14220 |
| GGCAGAGCAC  | CCCGAGTACG | TTAAGCAGCA | CGGACTGCGC | GTGGCGGTGG | ACCTGTACTT  | 14280 |
| CGACAAGCTG  | GTACACGCGG | TAGCCAACAT | CATCCAATGC | CTCTTCCAGA | ACAACACGTC  | 14340 |
| GGCAACCGTA  | GCTATGTTGT | ATAACTTTTT | AGACATTCCC | GTGACTTTTC | CCACGCCCTA  | 14400 |
| GTGACTCAGA  | CGCGGAAACA | GCGCTAGAAA | AGTTTCTCTT | TGCGCTATGT | GGGACAACTA  | 14460 |
| GAGTCCAACC  | TGGCAAGCAG | TGGAGCAAGA | CGCCAGACAG | CCGATCTCGA | AAAAATAAT   | 14520 |
| GCAGACAGAG  | GCAACGTTCA | TCCTAGGTGA | CTGGGAGATA | ACGGTGTCTA | ACTGCCGGTT  | 14580 |
| TACTTGCAGC  | AGCCTAACAT | GTGGCCCCCT | TTACAGATCT | AGCGGCGACT | ACACGCGGCT  | 14640 |
| AAGAATCCCC  | TTCTCTCTGG | ATCGACTAAT | ACGTGACCAT | GCCATCTTTG | GGCTAGTGCC  | 14700 |
| AAATATTGAG  | GATCTGTTAA | CCCATGGGTC | ATGCGTCGCG | GTAGTGGCCG | ACGCAACGCG  | 14760 |
| CACAGGCGGC  | AACGCGCGAC | GCATCGTCGC | GCCTGGCGTG | ATAAACAATT | TTTCAGAACC  | 14820 |
| CATCGGCATT  | TGGGTACGCG | GCCCTCCGCC | GCAAACGCGC | AAGGAAGCTA | TTAAGTTCTG  | 14880 |
| CATATTTTTT  | GTCAAGTCCC | TGCCCCCGCG | GGAGATGACC | ACATATGTGT | TCAAGGGCGG  | 14940 |
| CGATTTGCCT  | CCCGGAGCAG | AGGAACCCGA | AACACTACAC | TCCGCCGAGG | CACCCCTACC  | 15000 |
| GTGCGCGGAG  | ACGCTGGTAA | CTGGACAGCT | GCGATCCACC | TGCGCGCGAA | CGTATACGGG  | 15060 |
| ATACTTTTCA  | AGTCTGTGCC | CGCTCTCTTT | TTTGGACCTC | CTGACATTGG | AGTCCATTGG  | 15120 |

GTGTGACAAAC GTGGAAAGTGG ACCGCGAGCA ATTGACACCC AAGTACTTGA CGTTCAAGCA 15180  
 GACGGGAGAA AGACTTTTGA AATAACCGT TTACAACAC CATTCGACAG CATGCAAGAA 15240  
 GGGCCGTGTT CGTTTCGTCT ACAGACCGAC GCGGTCCGCG CGTCAGCTTG TCATGGGTCA 15300  
 GGCTTCACCC CTCATAACAA CCGCTCTGGG AGCCAGGGTA TTCGCACTCT ATCCAGACTG 15360  
 TGAGAAACT ATCCCACTTC AGGAAACCA CACCCTGAGG ATTCAATTGC TGTTCGAGCA 15420  
 GCATGGTGGC AAGCGCGGAG ACTGGCGCTT TGTCAATG GGGCTCGCCC GTGAACAAA 15480  
 GTTGTCTCA TTTCGCGCAG TACTCCTTCC GGGCAAGCAC GAACACCTTA TTGTATTCAA 15540  
 CCCACAGACA CATCCTCTGA CCATTCAACG GGACACAATA GTGGGCGTGG CAATGGCTTG 15600  
 CTATATCCAC CCGGGTAAGG CAGCCAGCCA GGCACCATAC AGCTTCTACG ACTGCAAGGA 15660  
 AGAGAGCTGG CAGGTGGGGC TCTTCCAGAT CAAACGCGGA CCGGGAGGGG TGTGTACACC 15720  
 ACCTTGCCAC GTAGCGATTA GGGCCGACCG CCACGAGGAA CCCATGCAAT CGTGACTGTC 15780  
 CGAGCACATA TGGCGCAGGA CTCAGAGCAG TGCTCCCGTG CTTTGCACT GTGCACTAGT 15840  
 AAACGACAGC TCGGGCGCGG CGAGCCCGTG TGGGATTCCG TCATTACCCC GAGCCACATC 15900  
 GTCATCTCTA ATCGAGTACC COTCTTACTA AGAGAACAGC ACATATGTCT CCGTTGCTGC 15960  
 CCCAGCGTGG GCCAGATCCT CCACAGAGCC TACCCCAACT TTACATTTGA CAACAGCAC 16020  
 CGCAAGCAGC AAACGGAGAC CTACACTGCA TTCTACGCTT TTGGGGACCA AATAACAAAG 16080  
 GTTAGGATCT TGCCCCACTGT TGTGGAAAGC TCCTCGAGCG TGCTGATTTT TAGACTGCGT 16140  
 GCATCGGTCT CTGCGAACAT CGCGGTGGGA GGGCTCAAAA TAATAACT TGCTCTCAC 16200  
 CTGGTGCAATG CCCAAGGAGT GTACCTGCGT TCGGGTAAGG ACCTTTCTAC ACCACACTGC 16260  
 GCACCGGCTA TTGTTACGCG TGAGGTGCTG AGCAGCGGGT TTGAGCCGCA GTTTACCGTA 16320  
 ACTGGCATTG CAGTGACATC CTCGAACCTA AACCAATGCT ACTTTCTGGT AAGAAAGCCA 16380  
 AAAAGCCGCG TGGCAAGCC GTTTGCAGCG CTGTCCGCGG AGACGACTGA GGAGTGTGCG 16440  
 GTCAGGTCTA TCCGCTTGG GAAGACACAC CTGCGGATAT CCGTGACTGC GCGTGGCGAG 16500  
 GAAACGCCCC TGTGGGGGCT CGTGACCACG AGCTTCAGCC TTACCCCCAC CGCACCGCTG 16560  
 GCGTTTGATC GTAACCCGTA CAATCAGGAG ACATTTGCGT GTAATGCCAA GCACTACATC 16620  
 CCAGTCATCT ACAGCGGACC AAAAATTACG CTGGCCCCGC GCGGCGGCA GGTAGTCTGG 16680  
 CACAACACCA GCTACACGTC CTCCTTCCCA TGCAAAAGTCA CAGCCATCGT CTCAAACCA 16740  
 TGCTGTAAC GTGACATATT TTTAGAGGAC TCGGAATGGC GCGCAACAA GCGAGCACCC 16800  
 CTGAAACTGG TGAACACGAG TGATCATCCC GTCATATTGG AGCCGGACAC ACACATTGGA 16860  
 AACGCCCTCT TCATCATCGC ACCCAAGGCC CGAGSTTTAC GCAGACTGAC TGGCTTAACC 16920  
 ACAAAAACCA TTGAACCTTC TGGCGGGGTA AAGATAGACA GCAGGAATT ACAACATTC 16980  
 AGAAAAATGT ATGTTGCCAC CGGACGCACT TAGGTGTCCG GTTCCCACCC ACACATTTGT 17040  
 CTTTATTGCT TTCAATAAA ACGGTGTTCT GTCAACCTCC TCCGGGCTCA CTAGTATTGT 17100  
 GTTCCCATAC GCGCTGTGCG CCCCAGGATC AACACTTCGT CCGCTATCCA CCTAATACA 17160

|            |             |            |            |             |            |       |
|------------|-------------|------------|------------|-------------|------------|-------|
| TAACACACAC | AAAGACATAG  | TGACTGTAGA | CAGTTAATCT | TTATTGTCTA  | GACACGCAAA | 17220 |
| GTATATTAGT | GTTATAAGAA  | ATTTTATGTC | ACGTGCGTCT | TTACTTATCG  | TGGACGTGAG | 17280 |
| GAGTCACGTC | TGGGATAGAG  | TCCAAAACAC | GCACCGGCTG | ACCTGCAAAAC | TTTTCCATTG | 17340 |
| CACTCAGAAC | ATAAAACGAA  | GCAAAGTGTC | TCACCCAATA | CTTAAGTCCC  | TGAAGGCTCC | 17400 |
| CTAATAGACC | GCGGTCAAAAT | TTGGGTGGAC | TGTAGTGCGT | CTTAGTCAGC  | TTATTGAGCT | 17460 |
| CTTCCTGTAT | GTCCCATCCT  | AAGGTCTTCG | TCAGAAGCTC | CATGACGTCC  | ACGTTTATCA | 17520 |
| CTGATTTTCC | AAACTCCGTC  | GTTAAAAACT | TAAACAACAC | CTCGAATTCA  | AAAAAGCCAT | 17580 |
| CGGCGAGCCT | TTTAAGGCAG  | CTAGTCTCAT | TAAATCCTAT | TAACCGGCAG  | TGATCASTAT | 17640 |
| CGTTGATGCC | TGGTAGTTTC  | AGATGAAAAA | TAGCAGCGGG | CTCTAGAATA  | CCCTTGCGAG | 17700 |
| TGCCGCTACG | GTAACAGAGG  | TGCGGGAAGC | ATTCATCGAT | CACCCATAGC  | ATCCAATTGA | 17760 |
| GTCTCTGAAT | GAGAAGATCC  | TTTTCAAAC  | CGGGGGCGTC | CGGCAACTTG  | CCCCGCGTTC | 17820 |
| CAGATACCAG | CAGTGAACCG  | ACCTAGCAAG | GAGACCACAA | CTTGAACCAG  | CACATGGCTG | 17880 |
| CTAACGCGGC | ATACACTAGC  | CGGTGGTGCC | CGAGCGGGAG | TTACGAAGTC  | TCACTGAAGG | 17940 |
| CGCGGCTCCT | GGGTGGGGGC  | CGCTCCAAAT | CAGGCAACGC | CGTATCCGAA  | CTCTGAGTCA | 18000 |
| CTTTTATGTA | GCTCTCAAA   | ATGTAAAGA  | TACCACGTTT | TTGAAAAACC  | CTCTCTTGCT | 18060 |
| CGCCAGGCTT | GGGCTTCACG  | CGGSCATACG | CAGCCAAGCT | ATCATGCGAG  | AGAAACACGT | 18120 |
| CACACGCAAA | GTCTGTATAA  | ACCGGGGTTA | AAAATAGCCT | AACTGGCCAG  | GGGCCAGTGA | 18180 |
| GGCCCTCCCG | GTACAGGTCC  | CCACCCCGCA | TGACCCAAAC | CTTGTCATT   | TGCTGTGCTA | 18240 |
| GCTCTGGGCT | TCTCGCCAAC  | CCAAGCGCGG | CATCGAGCGA | ACTCGCCAAA  | AAGTGAGCAC | 18300 |
| CAGGGGGCGG | GGTTTCTAAC  | GTGCGACTTA | GAACCACATT | GATTCTAGCC  | GCCAATGGTC | 18360 |
| GACAGCCCGC | GGGAATCGAA  | AGCCATGTGC | GCGCCCCCAT | AACAACCATG  | TTTTGTTTTT | 18420 |
| CAGGGGCACA | GTGGGTAGTC  | AGCTGTGCGA | AACGCTCAT  | GTCTCCCGGC  | AATGCAGGCG | 18480 |
| ACGGGAGACA | TCTGTTTTTT  | CCGATCCCGA | GTTTGGTATC | AACCGCAACT  | ACACAGTAAA | 18540 |
| GTGTAGGATC | CATGCCGCGA  | GGGTATAGGT | AAACACCACC | AACCACACAG  | TGTGCTCTTA | 18600 |
| TATACTTTTA | ATGAAACATA  | AGGCGAGACG | AAACAGCCGA | ACGTTTCTTA  | ATCAGGCCCA | 18660 |
| TGGAACCAT  | GCCACCCCA   | GGCAAACCT  | GTGGAAGGAT | ATCAACTAGA  | GAGGAGGGTC | 18720 |
| CAGCCTTATT | ATGGCAGGAG  | ACACTATAAG | CCCCATCGCC | CGACTGGGCA  | CCACATAAAC | 18780 |
| CGCCACAGTA | AGTGGCCCTA  | TACCGCTCAG | CGCCCAAGTT | GTTACAGTCA  | CACCCAAACG | 18840 |
| CGGTTGGCTC | TACATTGTCA  | TCACGTCCAT | CATTATGTGT | TGGTTCTCCC  | GCTTCCTTGT | 18900 |
| ACCCGTGAGC | TTCATCCACG  | GATTCTTCTG | AGTCGCGATG | CACAGGAGCG  | CCATCCGCGG | 18960 |
| GGCCATCTTG | GTGCGCTGGA  | GCTGCCCCCG | CGGGGCCATT | TTGGTCCGCT  | GGAGCTGCCC | 19020 |
| CCGCGGGCCC | CTCCTCGTCC  | TGTTATCCC  | CACGGGGAAG | AATTTCTTGA  | AGCTCGATCT | 19080 |
| CCTCTACTGC | ACACTCTGGT  | GATGTGCGCC | GAGGTCTATA | TGGAAACACT  | TCAACCCGCG | 19140 |
| TGTTTACAGC | AGCGTATGCC  | CGCCCTACGT | GGCGCATCAT | GTGGAAAAAC  | GCACCCAACC | 19200 |

|            |            |             |            |            |            |       |
|------------|------------|-------------|------------|------------|------------|-------|
| CAAAAACGAC | AAACAATTGG | TAAAACACGA  | AAAAACGTA  | GTACGGGGGT | GCAGCGACGT | 19380 |
| GATCTATCTC | TGGGTCTATG | CGGCCCCACTA | TATATAGCCA | AACCCACGTC | GCAGCGGCAA | 19381 |
| GGCCAGCGGC | CCCCAATGTC | ATAATGAAAA  | TAAAAACAT  | CAGTTCCAGA | CCCTCCTGGT | 19382 |
| AAGTCAGCCG | AGGCAATAGC | GTCAATTCGC  | GCAAGGGTGG | CCAGACCCAG | CGCTCTTGGT | 19440 |
| ATACGACGCC | ACATATCTGA | CAGGCGCTGT  | TTCTAGAGAT | AGTGAGCCAG | GTGCTTAAAC | 19500 |
| AACCTCTATG | GACGTTCTCG | AGCTCTCCTG  | TGCATCCACA | GGCTCTAAAT | CTCTCATTTG | 19560 |
| CGAGCTCCTC | GTTCGAATC  | CAGCAGACAG  | GAACATCCTC | ATCTTCCATA | TCCTGAGAGA | 19620 |
| GAACCCACAA | TAAAACATGG | CATTAACCCC  | TGCAACAAGT | GACCGTACCA | GGGCAAGCGT | 19680 |
| CCAGGCAACC | GGGGTCCCCC | TCGTTGGTCT  | ATACAATTCC | ATGACTACCT | ACTGGAATG  | 19740 |
| CTACAGCCAC | TCCTGTACA  | AGCCGGTTAA  | CTGGGAGGCG | ACGCTGGCGT | GGTATCGGCC | 19800 |
| AACCTGAACA | CACCACTCCA | CTCCAAACAC  | TTATGTACTT | TGTGGCTCGG | CTTTATTGTA | 19860 |
| ACAGCCAAAG | GGGGCGTTTG | TGGCTCAGCT  | TTATTGTAAC | AGCCAAGAGG | GACGTATGTG | 19920 |
| GCTATCTCAC | AAAAAGTCAC | CGATTCATGT  | AGACAACCCG | CTCCCACGAA | TTGGSTTTTT | 19980 |
| AAAAAGCCCT | CACGTATACA | GACGGGCCAC  | TAAATACGCA | CATGAGCCGG | CATCCTGTTT | 20040 |
| CGCGCTTGAC | GCCCACCACT | CTGACCGCAC  | GCTAAACATC | GCCCTACCTG | CTATACTGCC | 20100 |
| ATTTCCATAC | GAATGGTAGG | ATGCGGGCAG  | TAGTCCACCA | GTCTAAAATC | ATCAGGTGTA | 20160 |
| AACCTCTCCA | TGGAGAAAC  | AGACCGGAAT  | ATCTCCAGGC | GCGGAAAGGG | ACGTGGAGTG | 20220 |
| CGCGTCAGCT | GCAGCCGTAG | TGGCTCTATA  | TGCGTTTTGT | AGATGTGGGC | ATCTCCCAAC | 20280 |
| GTGTGAATAA | ACTCCCCGGG | TCTAAGACCA  | GTAACATGAG | CAAGCATATA | AGTTAAGAGG | 20340 |
| GAATAGCTGG | CAATGTTAAA | AGGAACTCCC  | AAACCCATGT | CTCCCGACCT | CTGATACAGC | 20400 |
| TGACAGGAAA | GCTCACCCTC | AGCTACATAA  | AATTGACATA | ACAAATGACA | GGGCGGAAGC | 20460 |
| GCCATCAACG | ACAAGTCCGC | CGGGTTCCAC  | GCACACATAA | TGATTCTTCT | ATCGTGCGGA | 20520 |
| TTATTTTTTA | TTAAATCCAC | AATGTACGAC  | AATTGGTCAA | ACCCCTGGCC | TGTATAGTCA | 20580 |
| GCTCCCGCT  | CCACGTACGC | CGCCCCAAG   | TGCTCCACT  | GGAAACCGTA | AACAGGTCCC | 20640 |
| AAATCCCCCT | CCCTTCTGTG | CGCCAGGCCG  | CGCCCGGCCA | GGAACTCCCT | GGAGCCATTT | 20700 |
| TTGTCCCAT  | TCTTGACTCC | TGTTCTTGAA  | AGCTCCCTGG | AGTCAGTACT | CCCTTCAGA  | 20760 |
| AACCAAGCA  | GCTCTTGCA  | TACGCCTCGC  | CAAAACACCC | GCTTTGTGGT | TAGTAAGGGA | 20820 |
| AAGTGGTCCC | GCAGACTATA | CCTGGCCTGC  | ATGCCAATA  | GAGAGAGGGT | GCCTATGCCG | 20880 |
| GTGCGGTGGA | GTGATCGCT  | GCCACGGCAC  | AAAATTTCCC | TCAACTGCCT | GAGATACTGA | 20940 |
| AGTTCTCTGT | GGGGCGTCTC | AGCCCCAGTT  | ACCTCATGCT | GAATCGAACA | AGGGTCAACC | 21000 |
| TGGGGGGCCA | AAGCCAAGAC | GCCAGGCTTT  | TGACAGAAGC | GAAACCCCT  | GGCACGGGAT | 21060 |
| AACTTTTTGG | CGACATACAA | GCTTAAAGGT  | ACAAACGGAA | ACATGATAGA | TCCTGGAAGT | 21120 |
| TTGTGAAGCC | CTGTGCCCCG | AGAGACACCC  | CTCAACTCGC | AGTGCTCGGA | GACCTACATG | 21180 |
| TATACTCAGG | CTCTTCTATA | AACCCCTCCC  | AAAAGTTTAT | AAAACACCGT | ACGTAATACA | 21240 |

|             |            |            |            |            |             |       |
|-------------|------------|------------|------------|------------|-------------|-------|
| CATTACTCAC  | AGTTCCCAAG | GTGACGCCCC | AACCCATGCA | CACGGGCGTG | ATCGATACCA  | 21300 |
| GAAAACATCA  | CAAGAACAAA | AAGTGTGTGT | GTGACATTCA | CATTTATTTT | TACAAGACAA  | 21360 |
| TTTTGTGCAG  | TAGAGTTGTG | CCTTCCGACA | CCCCGCGCCG | TTGCGTGTTC | TCTGTAAAT   | 21420 |
| GGGAGATCCC  | ACTCCTTGGC | AGGCACSTTT | CACGAAACGC | TCTTGTCTCG | CTGGCCTTAG  | 21480 |
| ACTTGTGGAC  | CCAACATGGG | TATCGTTAGA | GATCCGTGCG | GTAAATGGGC | AGCTGGCAAA  | 21540 |
| GCATTCTTCA  | GCGAGCAGTG | ACTGGTAATT | GCTGCATCAG | CTTCTTCACC | CAGTCTTTCC  | 21600 |
| ATTTGTCCGC  | ACACACCTGG | CGACCACGCT | TTGTCAAAAA | TATCACACCC | GGCTTCTGTC  | 21660 |
| ACAGTTGGGA  | GGTGGGCTAC | CAGCTGGACA | GAAGCACCTG | TGGTAATGGT | CTTTTCTGGT  | 21720 |
| AACCGAGACA  | GCACCTGTCC | GGTCTATGCC | AGGACGCTCC | CAGCGTGTCC | CCAGATTGCA  | 21780 |
| AACAAAGCAA  | GGCAGTCAGC | ACAGCGACGA | GCAGGATGCC | CTTGGTGTCC | ATAACTCCCC  | 21840 |
| TCGTGTGTCC  | TCGTGTAAAT | GCGAAACGGC | GATGTTAGGT | CAGGCGCGGT | AAACAGCTCA  | 21900 |
| ACTCGSTTCA  | AAACACGTAC | GTGATGTAGT | GCTGGTTCTA | CGACGCCTAC | CTGTAAACTC  | 21960 |
| CAGGATCCTC  | GGCTTTTATT | ACGAAGGCCA | ACACCCCAAA | AAATCCACGC | CCCCGTGACC  | 22020 |
| GCAGGGGCGG  | TTACTAACGA | CGGTTACAGG | TCCCTCCCGA | GCCACGCACC | TGCCATGTAA  | 22080 |
| CCTGCAAGGT  | AACCAGACAA | ACATCTAGGA | AGCGTAAATA | TCCCCAGSTA | GGAGAAGTAT  | 22140 |
| TGCATATCTC  | ACAGACTCAA | CACACACGGG | CCGTTACGCA | ACGGCTAGGG | GCATAACCCCT | 22200 |
| TTACCGGCGC  | GAAGCGCTAC | GCGCTTCGCG | AGAGGTATCT | CCGTGTGCTT | CTCCATCAGA  | 22260 |
| AGACGCGTGC  | GCCGCTTCGC | AGGCGACCCG | CATACCTTCC | GCCCCGAGTG | CGTTACAAAA  | 22320 |
| ATGACTGCCT  | TCTGGCGACA | ATACACGGTG | GACGTCCAGT | ACCACCCGCA | TATCAGCTTA  | 22380 |
| TCCGGTGGCA  | ATCTGGCACT | GGACAGGGAA | TTCTCGCAAC | AATCCGAGGC | CATGATGGTG  | 22440 |
| GCAGGACCGC  | TGGCCGCACA | TAGCTCAATC | ACGGCCACCC | AGAAGAGCAG | CCCCAAATGT  | 22500 |
| GCGCGCAACA  | CCCAGCACAT | GCTCCACATA | CATTTCTGGC | GCCACAACGA | TGATGCGCAA  | 22560 |
| AGGGGTGCAT  | TACCCATAAT | CCGAGCCTAG | TTATAAATTA | TTGAAGCCCC | GGCGACCAGG  | 22620 |
| GGTCCGCGCG  | CTTTTCCTCC | CCAAACGCGA | CGATAAAGAC | CAGCGTTGCC | AAATGTAACT  | 22680 |
| TATGTATAAC  | CCAAAATATT | GCGCATCGAT | AAGGTTTGCC | AAAACACCCG | AAAGTACACA  | 22740 |
| CACAAAAAAA  | CAGCAACAAG | ACGCTCACTA | GACATTCAAC | CCTTCCCCCA | CCCCCGAAAA  | 22800 |
| CAAAACAACCT | TGACACAGGG | GAAACACCAG | GGGCGGCGGA | GGTTGTCAAT | AGTGTCCAGT  | 22860 |
| ATTTCTTTAG  | ACGCGGGTTC | TTGGACCCGA | TGTCCCAGGT | CATTAAAGTC | TCAAATGGGA  | 22920 |
| TTAAAGGATC  | ATAGTTCCCA | GGTTTAATAC | TCCAAGCTAT | CCCAGAACAG | GACCCCGGCA  | 22980 |
| GAACCCCGCT  | TAACAGCACC | AAATCCACTT | GCGGTCCCAG | AAAAGGTGCG | CGAGGTGGCA  | 23040 |
| AGGTGACTGA  | AAAGGTCATA | GAGAGGACAC | CGGTCCCAT  | TCCCACGGTC | CAAAAATCCA  | 23100 |
| GCGCGCCCCA  | CCGGCTTTCC | GAGAACTTCG | GCAAAGCTAA | TTTGCATGCG | CTAATCCTTT  | 23160 |
| TATGTGCATA  | AATTATGTAG | ATGAGGAGTC | GCGCATGCGC | AGAAAAATTC | AGAGCGCCCC  | 23220 |
| GGTGACGGGG  | GTCACCTCCA | GGTCACGCGG | CTAGGTGGGA | CCGTGAGCGA | CTCGAAAAAT  | 23280 |

|             |             |             |             |            |             |       |
|-------------|-------------|-------------|-------------|------------|-------------|-------|
| TATAATTTTT  | GGCCATTTC   | TGGGCGCCGC  | CATCTTGAAT  | TTGCTAATCC | CCCATATCT   | 23340 |
| TGTGCCCCGC  | TCCCATTTGGT | CCGCGCGCCC  | GTCAATCAAA  | GTTTTCCGAG | CCGCCATTGG  | 23400 |
| CCCATCCGGC  | CGACCAATCC  | CGTTGAGGCT  | AGGCGACCGC  | GCCATTCCAT | TGGACGCCCC  | 23460 |
| AGCCCTCAAT  | CAAAATTCGGA | GGCCTCCCAT  | TGCCCCCTAT  | CCCTAGAACT | CCCAAGCTGA  | 23520 |
| TTGGCCCCGA  | GCGGGAACCA  | ATCAGCGATT  | AGAGTTTTGT  | TTTGATTTTT | CCATATATATA | 23580 |
| TATATATAAT  | CCTTTAATCC  | TAGCGCAGCT  | GAGTCATCGC  | AGCCCTTATT | CCAGTAGGTA  | 23640 |
| TACCCAGCTG  | GGTAATCCAG  | TAGGTATACC  | CAGGTGGGTG  | AACCCAGCTG | GSTATACCCA  | 23700 |
| GCTGCAATTC  | TATAATTAAA  | CAAGGTAGAA  | ACCAACGGGG  | TCTTCAGGTG | GATTTTCCGG  | 23760 |
| AAGCATTACC  | AAATAAGGCA  | ACCTCAGCTG  | GGAATAACCAG | CGGACTACCC | CCAACTGTAT  | 23820 |
| TCAACCTCTC  | TTTGTTTTTC  | GGAAGTATAT  | CCATTTATGG  | AAATCAGCTG | GSTCACTCTA  | 23880 |
| CTGGGTTATT  | CTTTATAATA  | GGGCCCCGATG | AGTCATGGGG  | TTGGGATTTT | TCTACTAGGT  | 23940 |
| CGTTTCGGTG  | GATGGGTGCC  | AGGATTATAG  | GGGCCCTGTC  | CACGGGGTTG | TTGGGTGGCG  | 24000 |
| GGGGGGGGGC  | TAGTGAGTCA  | CGGGCCTGGA  | ATCTCGCCTC  | TGGGTGGTTT | CGGTAGATGG  | 24060 |
| GGGCCGGGAG  | GATGGGGCCC  | CGCCCAACGC  | TGGCGCGCCC  | CAGAACATGG | GTGGCTAACC  | 24120 |
| CCTACATGGG  | CAGCTTGTCC  | TACGGTTACG  | CCCATTTGAG  | ACGGGTTAAC | CAACTGTTAC  | 24180 |
| ACCCCTTCCG  | CGGGAACGCT  | ATAAAAACGA  | GGGACAGCAG  | CCCCCCTTCG | CGCACTGCGC  | 24240 |
| GC CGCGCGGC | ACGTGGGACG  | GATCTCTTGG  | ATTTACCCCT  | AACGAGGAGC | CCCGGCAGCA  | 24300 |
| CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC | CCCGGCAGCA  | 24360 |
| CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC | CCCGGCAGCA  | 24420 |
| CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC | CCCGGCAGCA  | 24480 |
| CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC | CCCGGCAGCA  | 24540 |
| CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC | CCCGGCAGCA  | 24600 |
| CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC  | CCCGGCAGCA  | CCCCAGGAGC | CCCGGCAGCA  | 24660 |
| GGAGGGGGAT  | CCCGGCAGCA  | CACCTTCCCC  | GGAGGGGGAT  | CCCGGCAGCA | CACCTTCCCC  | 24720 |
| GGAGGGGGAT  | CCCGGCAGCA  | CACCTTCCCC  | GGAGGGGGAT  | CCCGGCAGCA | CACCTTCCCC  | 24780 |
| GGAGGGGGAT  | CCCGGCAGCA  | CACCTTCCCC  | GGAGGGGGAT  | CCCGGCAGCA | CACCTTCCCC  | 24840 |
| GGAGGGGGAT  | CCCGGCAGCA  | CACCTTCCCC  | GGAGGGGGAT  | CCCGGCAGCA | CACCTTCCCC  | 24900 |
| GGCAACAACC  | TGTTGCCATG  | TATGGCGATT  | TGTATCACTC  | ACAAGCACAC | AACCCCTGCT  | 24960 |
| AGTATTAATG  | GTGTTTAAAA  | CGTTCTACAC  | GTACGGCGGA  | CGGCATCCGT | CGCAAGCAGG  | 25020 |
| CGCATATAAC  | CCCCAAATGC  | ACCATGATGA  | GAAGCACAGC  | CACGCGTCAA | AAAACCTTTAA | 25080 |
| AAACATCGTT  | ATCCAAATGC  | ATTAAAAACC  | ACACCGAAAT  | TTACACAGGT | AGCACGTCAAC | 25140 |
| CGTGTTAGTG  | TCACCCACTG  | TACACAAGGC  | GTGTGCTATA  | TGTAGTATAG | GTATTTGATG  | 25200 |
| AGGCGGAAGC  | ATATCCCGCT  | TCCAGCGAAC  | GGAATAAGA   | ATCATCGGTT | CCAGCATTTA  | 25260 |
| TTCAAAGAGG  | GCACAGAGGA  | TTACATTST   | TTAGAGAGAG  | TTTTTCTTAG | TCACCATTC   | 25320 |



|            |             |             |            |             |            |       |
|------------|-------------|-------------|------------|-------------|------------|-------|
| ATACTTGGGG | AGTATTGGCC  | TACGATTTGG  | GCGACGTTTT | AGGCTGGTCT  | ATTCTCCCTC | 25380 |
| CACTTTTCCC | CGGCTATTCT  | GTCCCAGCAT  | AGGCTCTTGA | AATAAACAAAT | GTTTACCGAG | 25440 |
| TAAAGGTTTC | CACCTACCCCT | CATTTGTCTT  | TGCACCCATC | CCCCCTTTGC  | TTAATCACCC | 25500 |
| GAAACTAGA  | GGACACGGAT  | GGAAAACATA  | TGCGACGCGG | GTTGTTTGAA  | AGTCAACAGC | 25560 |
| TACTTGTTTT | TAATGAGGAC  | AGATTTGGGC  | ACAGGCCAGA | GGGTAAAGCC  | CTACGTGTGC | 25620 |
| GCGGGGGGGG | GGGTGTATAC  | GCTGCGAAAA  | CCTGCACGGT | GCATAACACC  | CAGGGCGTCA | 25680 |
| CGTCACATAT | CTCTGTGCAC  | CCAAGTGGTT  | GTTCAACCGT | TGTTTTTTGG  | ATGATTTTTT | 25740 |
| CGCACCGGCT | TTTTTGTGGG  | CGCGCATAGG  | TGGGTACGCG | CTGTCCCCCT  | AAGTCCCGCA | 25800 |
| CGGTCTTTCC | GGCCCCCTTC  | CGGCTCGTCT  | CCGGATGAAC | CGTCACGTTT  | TTTGTCTCCA | 25860 |
| GAGGCGACGT | CTCCTTCAGA  | TGACTCGTCC  | GTGGGCTCCT | CGTCCGTCCC  | GCCCCGCGGT | 25920 |
| CCGACAAGGA | CCGTCAATTC  | GATGTTATCT  | TGCTTCGCGG | TTGGCCGGCG  | CGGCCGTCCG | 25980 |
| TATGGCAGTA | CGGTACCCCG  | GGTGTATTTT  | GCCGCGTATA | ATGCCCTCAC  | AGTGCCACTT | 26040 |
| ACGCGGCATA | TGCTGCCAAA  | TGCAACACAA  | ATAAATATTT | GGTAAACCCC  | AAAGAAGCAG | 26100 |
| AGAAAACCGA | GCACGGCCCC  | GGGGGAGAAAT | GTTCCCGCAG | GAGCAGTTAG  | GATGACCAGG | 26160 |
| AGCGTCCAGG | TGCACAACGC  | CACGCCGACA  | AGCCCAGCCA | CCACCACAGA  | CATCAGCAGA | 26220 |
| AACAGTTCAA | AAATTTCTTG  | GCGCTCCATC  | TCCGGCCACA | GGTTAAGGCG  | ACTACGCCAC | 26280 |
| TGCCTCCCGG | TGCGGTATAT  | AACGCGACAC  | ATTTGACAGG | CCGTGTTTTG  | AGACACTGTT | 26340 |
| AGCCAAGTGC | TTAAACACTG  | CGGCTGGACG  | ACATCCAGCT | CTCCGGTACA  | GGCGCAGGGG | 26400 |
| TGTATGCCCT | CGTTCCCCAC  | CTTTTCCCTA  | CATATCCAGC | AGATGGGTCC  | CTCTACACCC | 26460 |
| TCTTCTACGT | CCTTAGACGC  | CATCTCTGCA  | GCTGGGGTGG | AAGTCTGAAA  | AAGGGAAAGG | 26520 |
| GGAGGTGAGC | AGAGTGCCCA  | GTTAGTCTCC  | GACCCGCCGT | CCGCCCTACT  | GTCGCTATCC | 26580 |
| CGCCTTGACA | GATCTCTAAC  | GTATTCACGG  | ACGCCACATG | TGTGTCTATT  | TTCTACATTC | 26640 |
| CAGGCTTTCC | CTGGAAACTI  | GTCACAACCC  | ACCCTGCTTT | AGCTCTACAT  | CTGTATTTTT | 26700 |
| GTTTACGCAC | AGGATCAACG  | CTTCGTGCCC  | GTCCACCCCC | GCGCTCTCCG  | CCTGTGTTTG | 26760 |
| GAGGTTTTAT | GAGTGGTTAG  | TTCTAGGCAG  | CTCCGGACAA | GTTGTCCAAA  | ACACGGCGCG | 26820 |
| CCCCGCCCTT | CCTTCCCTCC  | GGATCCGCCC  | ACACCGGACC | TATGAAATAA  | GGGACACGCG | 26880 |
| TCATCACTAG | TTATGAGAGA  | AAAACCACAA  | CAGCTTTATT | GGAAAACACC  | TGAGTGGATC | 26940 |
| CCCCACCCCC | CGCGTACGAC  | AGGCGTTTTCT | GTGGTGCCTT | TCTGGGAAAA  | ACGTTTTTTT | 27000 |
| CCCATTTCTT | CCTCGACAGG  | TCTTCTAAGG  | TAGATAAATC | CCCCCCCCTT  | GCGGCTCTCC | 27060 |
| TAGAATGGCC | TAGGCGCAGG  | ATGGCGTTGT  | CGCCTCGAGC | AGTTGGGCGG  | CAGTGATATC | 27120 |
| TTCAACTTTT | GACCGTCTAA  | GCTATGGCAG  | GCAGCCGCTG | CATCAGCTGC  | CTAACCCAGT | 27180 |
| TTTTGGAAGG | GTCTGCGCAG  | ATCTGACGCC  | CTCGCTTGGT | CAGCAAAATA  | ACTCCGGGTT | 27240 |
| TTGGGCACGC | TGGGGACGTG  | GGATACCACT  | CTTTTAGAAT | TTGGACGGGC  | GGTGGGTGCT | 27300 |
| GCTGGAACCC | GTAGCAGCAG  | CTATTAGGCG  | TGTACGACAC | GAGTGACCCC  | GCGCTTTCTG | 27360 |

|            |             |             |            |            |             |       |
|------------|-------------|-------------|------------|------------|-------------|-------|
| TGGGCGTCAG | GTAAACCGTG  | GCAACAGTA   | CGCTAACGCA | GCATAAAACG | TGGACGCGGG  | 27420 |
| CCATCTGAG  | GTGCCAAGTT  | CGCAACAGTC  | TAAAGAAAAC | CGTAAAGGCT | ATTTGCGGTT  | 27480 |
| TCTGTTCTGT | CAGATGTAAC  | GCCGAGTTCC  | TTATATGCTT | ACCTGATTCT | GCTCTCACCT  | 27540 |
| GTTTATTTAT | AGTGGCGTAT  | GCTAACCGCT  | AGCTTACATG | CGGGATAAGT | TGGCCTAACT  | 27600 |
| CACCAAAAAC | GGGTTGCAGA  | CAAAAGTGAT  | TGTTGGGGCG | CTTACTTAGA | AGGTGTGAGG  | 27660 |
| GTTTCTAAGA | AACCCCGCCA  | ACGCCCCGAA  | ACCGCATGCG | TTCCAGTCGG | TGCGGCGCTGC | 27720 |
| GCCGGCGTCC | CTGTGGGGCC  | TTTGTTGGGT  | TTGAGTTCTG | TCATTAAGCC | AGGTTTCCAT  | 27780 |
| TGCCACCCGG | GCGAAAACAA  | GCCGGGTAGT  | TTGAGGGTGC | ATCTGGCGAT | CAGTGTACCA  | 27840 |
| TATTCCCAAG | ACCCATCAAC  | ACCGGTGCTT  | GAGGCGTGTG | TCTGTATGTG | TCACCGGAGA  | 27900 |
| CTGCATGTAT | CGTGCAATAT  | TGTATTGTGC  | GCTTGCGCGG | AGACAACATA | CGGACGACCA  | 27960 |
| AGTCAGGGGT | CACCTCCAGT  | GCACGCGGCT  | AGGTGGGACC | GTGGGCGAGC | CGAATAAATT  | 28020 |
| ATATATTTTT | TTGGCACGGT  | TGTGAGCAAC  | GCCATCGTGA | GTTGGTTAAT | ACCCCTCTAA  | 28080 |
| CGCATAGTCT | TTTTTTATTT  | GTCAACCAAC  | CAGTCAATCA | CCTGTCAATG | CGGCTCAGAA  | 28140 |
| GCACACGTCT | TGGGCCAATG  | CCGTGTTGGC  | GGGTTTGACC | ACGGTTACTG | ATAGGTAGAC  | 28200 |
| GASTCCGACA | ATCACACACG  | TCCGCCAGCG  | ATTTGCAGCG | CAGCTAAAAT | CGCGTGGCGG  | 28260 |
| GSTTGGTAGA | AGCAAAATTAT | CCAATGGTCC  | TGTTTGGGTT | TGTTTTGGGG | TTATCTACAT  | 28320 |
| ATTATATTCC | TTATCCCGAC  | TGGTTGCGGA  | AGTATTCGCA | GCTTGGGTAC | TCTGCTCGAT  | 28380 |
| TACCCCGTGA | ATAACTGGGC  | GGGGGGTGAC  | CCAACATAGT | GATTCGGTAG | ATTTGGGGGA  | 28440 |
| CTGGATGAAC | ATTAATGAAA  | GTTTATTAAAT | GTTTATCCGT | ATTGTGTATA | TGTAATTTGG  | 28500 |
| TTTCCATATT | TGGTAGGAGT  | ATGGAGTTTT  | CTTATGGATT | ATTAAGGGTC | AGCTTGAAGG  | 28560 |
| ATGATGTTAA | TGACATAAAG  | GGGCGTGGCT  | TCCAAAATG  | GSTGGCTAAC | CTGTCCAAAA  | 28620 |
| TATGGGAACA | CTGGAGATAA  | AAGGGGCCAG  | CTTGAGTCAG | TTTAGCACTG | GGACTGCCCA  | 28680 |
| GTACCTTTGG | CTGCCGCTTC  | ACCTATGGAT  | TTTGTGCTCG | CTGCTTGCCT | TCTTCCCGCT  | 28740 |
| TCTGGTTTTC | ATTGGTGCCG  | CCGATTGTGG  | GTTGATTGCG | TGGCTTTTGG | CAATATACCC  | 28800 |
| ATCCTGGCTT | TGGGCTAGGT  | TTTCCGTCTT  | ACTTTTCCCA | CATTGGCCTG | AGAGCTGTAG  | 28860 |
| TACAAAAAAC | ACCGCGCGGT  | CTGGAGCTCT  | CCATAAGCCC | GCAGAACAAA | AGCTGCGATT  | 28920 |
| TGCCCCAAAA | CCTTGCCATG  | GCAACTATAC  | AGTCACCCCT | TGCGGGTTAT | TGCATTGGAT  | 28980 |
| TCAATCTCCA | GGCCAGTTGT  | AGCCCCCTTT  | TATGATATGC | GAGGATACTT | AACGTGTCTG  | 29040 |
| AATGTGGAAT | ATAATGTGAA  | AGGAAAGCAG  | CGCCCACTGG | TGTATCAGAA | CAGTGGTSCA  | 29100 |
| CTACCTATCT | GCTCATTCTG  | TGTTTCGGTT  | CTGTGTTTGT | CTGATTCTTA | GATAGTGTTC  | 29160 |
| AGGTAATTCT | AGAAAGCGGA  | TTGAGTGTA   | ATCGGGCCAC | TTTGCCCTAA | ATGTGACAAT  | 29220 |
| CTGGATGTGT | ATCTTATTGG  | TGCGTTGTGA  | AGCATTTTAA | AATGCGTTTT | AGATTGTATC  | 29280 |
| AGGCTAGTGC | TGTAATGGTG  | TGTTTATTTT  | TCCAGTGTAA | GCAAGTCGAT | TTGAATGACA  | 29340 |
| TAGGCGACAA | AGTGAGGTGG  | CATTTGTCTG  | AAGTTTCAAA | GTGCTGTAAG | AACATTGGAC  | 29400 |

|            |            |             |            |            |            |       |
|------------|------------|-------------|------------|------------|------------|-------|
| TAAAGTGGTG | TGCGGCAGCT | GGGAGCGCTC  | TTTCAATGTT | AATGTTTTAA | TGTGTATGTT | 29480 |
| GTGTTGGAAG | TTCCAGGCTA | ATATTTGATG  | TTTTGCTAGG | TTGACTAAGG | ATGTTTTGTT | 29520 |
| GTAGGTGAAA | GCGTTGTGTA | ACAATGATAA  | CGGTGTTTTG | GCTGGSTTTT | TCCTTGTTCG | 29560 |
| CACCGGACAC | CTCCAGTGAC | CAGACGGCAA  | GGTTTTTATC | CCAGTGTATA | TTGGAAAAAC | 29600 |
| ATGTTATACT | TTTGACAATT | TAACGTGCCT  | AGAGCTCAAA | TTAAACTAAT | ACCATAACGT | 29700 |
| AATGCAACTT | ACAACATAAA | TAAAGGTCAA  | TGTTTAATCC | ATATTTGCTG | ACTTGTGTCT | 29760 |
| TGACTTGCTT | CGATTGGGAT | GGGGGTGTGG  | GATGGGGGTT | TGGGATGGGG | GTGTGGGATG | 29820 |
| GGGGTGTGGG | ATGGGGGTGT | GGGATGGGGG  | TGTGGGATGG | GGGTGTGGGA | TGGGGGTGTG | 29880 |
| GGATGGGGGT | GTGGGATGGG | GGTGTGGGAT  | GGGGGTGTGG | GATGGGGGTT | TGGGATGGGG | 29940 |
| GTAAATGACA | ATGGGGGTAA | ATGACAATGG  | GGCGTTTGGT | GACACATTTG | CCCCACCGTC | 30000 |
| GCCTGCCCGG | AACCAGCTTG | GTGATGTGCT  | GTCTGGCTCT | CAGGTGCACT | TTATGCCAAG | 30060 |
| CAGTTGAGGC | GCATTAGATA | TATAAACTT   | GGGTACACAC | CCTTGGTGCT | GTGCGCGTGC | 30120 |
| TATGTGCCCT | GSTGACCGTC | CACAATGGAC  | GAGGACGTTT | TGCCTGGAGA | GGTGTGSCC  | 30180 |
| ATTGAAGGGA | TATTCATGGC | CTGTGGATTA  | AACGAACCTG | AGTACCTGTA | CCATCCTTTG | 30240 |
| CTCAGCCCTA | TTAAGCTATA | CATCACAGGC  | TTAATGCGAG | ACAAGGAGTC | TTTATTGAGG | 30300 |
| GCCATGTTGG | CTAATGTGAG | ATTTCCAGGC  | ACCACCGGTA | TAAACCAGCT | TGGGTTGAGC | 30360 |
| ATGCTGCAGG | TTAGCGGGCA | TGGAAACATG  | AACTGGGGGC | GAGCCCTGGC | TATACTGACC | 30420 |
| TTTGGCAGTT | TTGTGGCCCA | GAAGTTATCC  | AACGAACCTC | ACCTGCGAGA | CTTTGCTTTG | 30480 |
| GCCGTTTTAC | CTGTATATGC | GTATGAAGCA  | ATCGGACCCC | AGTGGTTTTG | CGCTCGCGGA | 30540 |
| GGCTGGCGAG | GCCTGAAGGC | GTATTGTACA  | CAGGTGCTTA | CCAGAAGAAG | GGGACGGAGA | 30600 |
| ATGACAGCGC | TATTGGGAAG | CATTGCATTA  | TTGGCCACTA | TATTGGCAGC | GGTCGCGATG | 30660 |
| AGCAGGAGAT | AACGCGTAAT | TCGAGGTCCC  | CGGAAGAGTA | GAGGCTTGCA | TGTTATACAA | 30720 |
| ACAACATAAA | CATTAAATGA | ACATTGTCCA  | AAACGTATGT | TTATTTTTTT | TCAAACAGGG | 30780 |
| GAGTAGGGTA | GGAAGGGTAC | GTCTAATACG  | TAAGTGTTCG | CTACTGCTTG | TTGAGGAGCT | 30840 |
| CCTCGCAGAA | CATCTTGCGA | ATTTTAGATT  | TTGGACTAGA | GCGACTGCTG | GCTTCAACGC | 30900 |
| GGTTCGATGT | AGGGTTTCGG | GTAGGAGCGT  | CTTTCTCCAC | CGCCGCGCAT | GSTGTATGCG | 30960 |
| TGCTCTCCGG | TGCCTGTTGT | TGGATGCTCT  | GCGTCTGGA  | GGCGGGGGTG | GGTTCAGCGG | 31020 |
| GTGCTGCGCC | AACTACCGCG | AGTCCTGTAG  | AGACTGCGCG | GTGGCTCACA | TGTGGCTGAG | 31080 |
| CAAAAAGGAT | GGGCGCCGCT | TGCTGGAACCT | GACCGTGTGG | CGCCTGCACG | TAAATGGGTG | 31140 |
| GSTGTACGTA | GGTTCCCTCC | TGCTCCTTCA  | TTGTGCGGAA | TTGACACGGG | ACCGCTGAAT | 31200 |
| TGGCGTGGGG | CCTGTAGTGT | GGATCTACTG  | CGGCTGCTGC | TGCAGAGGAG | GACGGCGGGT | 31260 |
| GCCCTGCGTG | CCAACCGTTC | AGTTTCATCT  | CTTTGAGTTC | AGACTGTATT | TCCGCTATGT | 31320 |
| TCTTTGACAT | GGACAAGATA | TCCTTGTGAT  | ACGCCGGCTC | CTCTCCTGGA | AAGAGGTGTC | 31380 |
| CTTCGTGCTC | CTCTGCGCCG | CGCTTGCGCT  | TCCCCGTCTT | ATATCCAGGC | AGCTGTGGCG | 31440 |

|            |            |             |            |            |            |       |
|------------|------------|-------------|------------|------------|------------|-------|
| AGTAATACCA | TGGATCGTAT | GGGTCTTTGT  | AAGCGTAGCC | GTATGCTGCG | GCTGGGTTTG | 31501 |
| AAACATACCA | AGGTAGGTGA | TGGTCGGTGG  | GGACATCTG  | GGCCCAACAC | GGCATTAGGC | 31560 |
| CTGGCCCTGA | AAGTGTATGT | GACATTTTGG  | CGGCTGTGGT | CTTCATTCCA | TGGATGCTGC | 31621 |
| TTTGTAGCAT | GCTCAGGAAG | GCGGATTTGG  | GGATGGATAT | GATATCCTCT | TGACCAGAGC | 31680 |
| TGTTCATGGC | TGGTCTGGGT | GCTGTGACGG  | CTTGGATGCG | GACCGGGAAT | TGGCTGGCCT | 31740 |
| TTAAATACCG | CGGGCTCAAT | ATGCTGGCCA  | CACCTCTGTC | AGTTTTCAAT | AGGTGAGGGC | 31800 |
| GCTCCCGTAT | GAAGCTGGCA | TCTATAGCTT  | TTGCCATTAA | GCTCTCCAGG | GGACTGACGA | 31860 |
| AATTTGGTGT | GGAAAGGTCC | TCCAGCCTGC  | AGCTACTTAC | GTGCTGGAGG | ATGTGGGGCG | 31920 |
| GCTCCGACTT | AGATACTGAT | GAGAATCTGG  | AAACCACCCA | CTCGGCGTCC | TGTCCGTACA | 31980 |
| CGGCCACTGT | GGCGCTCGG  | CGCCCCAGGG  | CGCATAGTGA | TACGTCTTGA | AACACGGGAC | 32040 |
| CGCTGGGAST | CTGGGATAAC | TGCGGGGGAT  | GTATAGACGA | TAAAGACAGC | CCCCGGAGCC | 32100 |
| ACGTGTGGAG | TATCTCCAAC | AGTGGTTCCCT | TAGGGAGATT | TTTCACGGGG | GCTCTGGCCA | 32160 |
| CTGGGGAGGT | GTCCGCCAGC | CTGGATGCCA  | GCTCTAGGAA | GGCTGGCGAC | GTGATGGCTC | 32220 |
| CGGTGCAGAA | AATACCTTGG | GACACTTGAA  | ATAGACCCAG | TGTCCAGCCC | ACTTCTGTCT | 32280 |
| CTGGTAGGTG | TTCGATTGTT | ATTGGAAGGG  | CTTCTGTGAC | TGGGAGATAA | TCCGTACCTT | 32340 |
| GATCCGGATC | GAGATAGAGC | TCTTGCTCCA  | GCTTGGGGCA | GGACACAACA | TCTACAAACC | 32400 |
| CTCCGACGTA | CAGGCCCTGT | GCCATGCTCG  | GAAAATACGT | GTGTGAGACC | GAGCCGCTGA | 32460 |
| GGCCGGGGCT | TAGGAGGCTC | ATGTGGCGCT  | TTTTGCCAAA | TAAGAATTTA | AATACATTCC | 32520 |
| ACGCCCAAGA | GCTGCGTTTT | ATTGATTTGG  | TTCTCTGCAG | GATGTACAAT | TTCGGTCTAA | 32580 |
| ATGTGTACCT | GTTAAGGGAG | GCTACTGCCA  | ATGCCGGGAC | CTACGACGAG | GTGGTCTCTG | 32640 |
| GACGCAAGGT | TCTGCGGAG  | GTGTGGAAGC  | TGCTGTACGA | TGGGCTCGAG | GAGATGGGCG | 32700 |
| TGTCAAGTGA | GATGCTGCTG | TGTGAGGCAT  | ACCGGGACAG | CCTCTGATG  | CACCTGAACC | 32760 |
| ATAAGGTGGG | GCTCTTGAGG | GGCTTGGCGA  | ATTATCTGTT | TCACCGGCTA | GGGGTCACCC | 32820 |
| ACGACGTTCC | CATCGCCCCG | GAAAACCTGG  | TGGACGGAAA | CTTTTGTGTT | AATCTGGGAA | 32880 |
| GTGTGCTCCC | CTGCAGGCTG | CTCCTTGCGG  | CGGGCTACTG | CCTCGCCTTT | TGGGGCAGCG | 32940 |
| ATGAACACGA | ACGCTGGGTG | CGCTTCTTCG  | CCCAGAAGCT | TTTCATTTGC | TACCTGATAG | 33000 |
| TCTCCGGGGC | TCTTATGCCA | CAGAGGTCTC  | TGCTAGTTTG | GGCCAGCGAA | ACGGGCTATC | 33060 |
| CCGCTCCGGT | GGAGGCAGTC | TGTGCGGACA  | TCCGCTCCAT | GTACGGCATA | CGAACGTATG | 33120 |
| CGGTCTCGGG | TTATCTTCCG | GCTCCGTCCG  | AAGCGCAGCT | GGCCTACCTT | GCTGCTTTTA | 33180 |
| ACAACAACGC | GGTTTAAACG | ACCGCGAGGA  | CCACCGGCAG | GCAGCCAAAG | ACCATAAAGT | 33240 |
| ACGCTCTATC | GTAGTCATCG | CCGCCGCCAA  | ACTGGGACTT | GATAATCTCC | TGGAGAAGGG | 33300 |
| TGGGTGGGGA | TGGGTGTGAA | ASCAGGACGT  | CCAGGCCCTC | TTCTGTTGCC | AGGCGGAGGG | 33360 |
| CTGTTCTCGC | CTGGAGCAGC | GCCAGTGGAT  | CTCGGAATGT | AAGCTGCTGG | TTGAGGATTT | 33420 |
| CGAATATCTC | ATTAAACCTA | CTGCCTGTCA  | GATTTACAAA | TGGTCCGGGT | TGTTTGTGGG | 33480 |

```

ACACGGTCCA TCGCGCCTCG AGGGCGGGCCA GTATTATGCC AGGGAAGATG AAGGACACGG 33540
GGGCGTTTGG ATTAGCCTGC AGTGTGGGGA TTATGTAGTG CTCGATATG AACGAAAATA 33600
GCTGGCCCCCT TTTGAGCATG GGGCGCTTTG GATCCGGTAG GGCACCGGGC TGAATTTTGG 33660
GTCCCAGCAG GATACACAGG TTCAAGCGGC GGTTTGGGTG CCTCGCGCG ACTTGCCCAA 33720
ACTCCAGCAA TCCATACGCG AGGATAAACA CCTCCAGCGC AACAAATCCC GCTCGCAGGT 33780
TCCACTGGTA TCGCGAAAAT GGTGGTATAT CGGACCCAAA CATGGCGCTC GTAATGGCGA 33840
ATACCAAGTC CATGGCGGGC GCTGTCCCTG GCGCGCCCGT ACCCTTGTGT TGGGGAAAATA 33900
ATCCAGCCTT AGCCATCATT GCGTGAAGCT TGTGGCGCTG GAAGAAGGCT GTCGGATAGC 33960
GGCTCTCCTT ATTGAGAGGC GCCAGCGAGG CGCGCTCCTG GGGGTTTGAG TATGTGAAGC 34020
TGAAGTCCCC AGGACCCTT TCCTGTTTTA GCTGAGTGAT TAGCAGGTCT AGCTTTTGAG 34080
GCAGGTCTGC TAACAGGTCA TCGGGAGTAG CGGGCAGTTG CCTGGATGTC TTTTGACAAA 34140
AGTACGCCTT GACGAGGCAA AGCGCGGCCT GGGTGTCCGT GAGATGCCTG GCGTCGGCGA 34200
AAAAGTCAGC GGTGGTGGAG GCGACCCTCG TCAGGGTGTG AGAGATGAGT TTGAGCGATG 34260
TGGAATTCTG AAAGTTAACA GTCCCTTTTA GTTCTTTAGG GAAGACGCGC CGCTGCATGG 34320
CGTTGTCCCT GAGGCTGATG AACCACGGCC CAAAGGATGG CAACCACTGA TTCTGTTTCA 34380
TGTACAGGCT GGGCATGAGC TCGCCGCGCA GGTCCCTGTC AACGGAGAAG TGAGGGTCCC 34440
CGGGGACGAT CGCCACGGTG AAGTTACGGT GGCTGGCCTG CCGGGGGGAT GTCAC TAAGG 34500
GAGGCTCATG GGAACGGCTT TGGGGCATGT CTATGTTGTC AGACCATGTC ATGTTGCCTA 34560
TCATCTGTTT CACCGCGTCG ATATCTGCGT TAATGACGCG GACGCGTGAG TCATGACCT 34620
GAACAAGCCG GTCCAGCTCT AGGGAAAGCA GGTGTGCCTT TGTCTTTTCT TCTCGATTTC 34680
GCACGAGTTG GCTGCGCAGT CCAAGGGCGA CCCTTCTTGT TTCTTCCATG GTGGGCTTGT 34740
GAATAAACAG CACGTTTTTC GGGTGTGGGG CCCAGAATCT TCCCGCCTCT GTCCATCTTC 34800
GGTTTTTTGG GTACCTTAGA TAGGACCTTT CTGATGTCAG CATTTTCTCT AGCAGTGAGA 34860
AAGGCGCACA ATTTTCCTTC GGTGGTGTGC ACCGGCGTGG GAAACGCCCC GGGTGATTCA 34920
GAGTATACTG TCTTTAGTGT TTTCTGATTC TTAAATATCA GCAGGGGCGT GATAGTCCAC 34980
GCCTCGGTAC CCGGAGGGGC CGAGTGAGCG ATGTAATGGA TCGAGTCGGA GAGTTGGCAC 35040
AGGCCTTGAG CTCGCTGTGA CGTTCTCAG GTGTTGGTTG GGATCAGCTG GTGACTCAGA 35100

```

## (2) INFORMATION FOR SEQ ID NO:18:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35100 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:18:

|                        |                       |                        |      |
|------------------------|-----------------------|------------------------|------|
| CAAGTCTTGA GCTCTACAA   | GTAACATACG GGCTGATGCC | CACCCGATAC CAGAATTACG  | 60   |
| CAGTCGGCAA TTCTGTGCC   | TAGAGTCACC TCAAAGAATA | ATCTGTGGTG TCCAAGGGGA  | 120  |
| GGGTCTCTGGG GCGGGCTACT | TAGAAACCGC CATAGATCGG | GCAGGGTGGA GTACTTGAGG  | 180  |
| AGCCGGCGGT AGGTGGCCAG  | GTGGGCCCGG TTACCTGCTC | TTTTGGCTGC TGCTGGAAAGC | 240  |
| CTGCTCAGGG ATTTCTTAAC  | CTCGGCCTCG GTTGGACGTA | CCATGGCAGA AGGCGSTTTT  | 300  |
| GGAGCGGACT CGGTGGGGCG  | CGGCGGAGAA AAGGCCTCTG | TGACTAGGGG AGGCAGGTGG  | 360  |
| GACTTGGGGA GCTCGGACGA  | CGAATCAAGC ACCTCCACAA | CCAGCACGGA TATGGACGAC  | 420  |
| CTCCCTGAGG AGAGGAAACC  | ACTAACGGGA AAGTCTGTAA | AAACCTCGTA CATATACGAC  | 480  |
| GTGCCACCCG TCCCGACTAG  | CAAGCCGTGG CATTTAATGC | ACGACAACTC CCTCTACGCA  | 540  |
| ACGCCTAGGT TTCCGCCCAG  | ACCTCTCATA CGGCACCCTT | CCGAAAAGG CAGCATTTTT   | 600  |
| GCCAGTCGGT TGTCAACGAC  | TGACGACGAC TCGGGAGACT | ACGCGCCAA TGGATCGCTTC  | 660  |
| GCCTTCCAGA GCGCCAGGGT  | GTGTGCTCGC CCTCCCTCTC | CGCCTCCAAA TCACCCACCT  | 720  |
| CCGGCAACTA GCGCGGCAGA  | CGCGTCAATG GGGGACGTGG | GCTGGGCGGA TCTGCAGGGA  | 780  |
| CTCAAGAGGA CCCCCAAGGG  | ATTTTTAAAA ACATCTACCA | AGGGGGGCAG TCTCAAAGCC  | 840  |
| CGTGGACCGG ATGTAGSTGA  | CCGTCTCAGG GACGGCGGCT | TTGCCTTTAG TCCTAGGGGC  | 900  |
| GTGAATCTG CCATAGGGCA   | AAACATTAAA TCATGGTTGG | GGATCGGAGA ATCATCGGCG  | 960  |
| ACTGCTGTCC CCGTCACCAC  | GCAGCTTATG GTACCGGTGC | ACCTCATTAG AACGCCTGTG  | 1020 |
| ACCGTGGACT ACAGGAATGT  | TTATTTGCTT TACTTAGAGG | GGGTAATGGG TGTGGGCAA   | 1080 |
| TCAACGCTGG TCAACGCCGT  | GTGCGGGATC TTGCCCCAGG | AGAGAGTGAC AAGTTTTCCC  | 1140 |
| GAGCCCATGG TGTACTGGAC  | GAGGGCATTT ACAGATTGTT | ACAAGGAAAT TTCCACCTG   | 1200 |
| ATGAAGTCTG GTAAGGCGGG  | AGACCCGCTG ACGTCTGCCA | AAATATACTC ATGCCAAAC   | 1260 |
| AAGTTTTCTG TCCCCCTCCG  | GACGAACGCC ACCGCTATCC | TGCGAATGAT GCAGCCCTGG  | 1320 |
| AACGTTGGGG GTGGGTCTGG  | GAGGGGCACT CACTGGTGCG | TCTTTGATAG GCATCTCCTC  | 1380 |
| TCCCCAGCAG TGGTGTTCCT  | TCTCATGCAC CTGAAGCAGC | GCCGCCTATC TTTTGATCAC  | 1440 |
| TTCTTTCAAT TACTTTCCAT  | CTTTAGAGCC ACAGAAGGCG | ACGTGGTCCG CATTCTCACC  | 1500 |
| CTCTCCAGCG CCGAGTCGTT  | GCGGCGGGTC AGGGCGAGGG | GAAGAAAGAA CGACGGGACG  | 1560 |
| GTGGAGCAAA ACTACATCAG  | AGAATTGGCG TGGGCTTATC | ACGCCGTGTA CTGTTTCATG  | 1620 |
| ATCATGTTGC AGTACATCAC  | TGTGGAGCAG ATGGTACAA  | CATGCGTACA AACCACAAAT  | 1680 |
| ATTCCGCAAA TGTGCTTCCG  | CAGCGTGCGC CTGGCACACA | AGGAGGAAAC TTTGAAAAAC  | 1740 |
| CTTCACGAGC AGAGCATGCT  | ACCTATGATC ACCGCTGTAC | TGGATCCCGT GAGACATCAT  | 1800 |
| CCCGTCGTGA TCGAGCTTTG  | CTTTTGTTTC TTCACAGAGC | TGAGAAATT ACAATTTATC   | 1860 |
| GTAGCCGACG CGGATAAGTT  | CCACGACGAC GTATGCGGCC | TGTGGACCGA AATCTACAGG  | 1920 |
| CAGATCCTGT CCAATCCGGC  | TATTAACCC AGGGCCATCA  | ACTGCCCAGC ATTAGAGAGC  | 1980 |

|             |             |            |             |            |             |      |
|-------------|-------------|------------|-------------|------------|-------------|------|
| CAGTCTAAAG  | CAGTTAATCA  | CCTAGAGGAG | ACATGCAGGG  | TCTAGCCTTC | TTGGCGGGCC  | 2340 |
| TTGCATGCTG  | GGGATGCATA  | TGGTTGACAT | GTGGAGCCAC  | TGGCGCGTTG | CGGACAAACGG | 2100 |
| CGACGACAAAT | AACCCGCTCC  | GCCACGCAGC | TCATCAATGG  | GAGAACCAAC | CTCTCCATAG  | 2160 |
| AACTGGAATT  | CAACGGCACT  | AGTTTTTTTC | TAAATTGGCA  | AAATCTGTTG | AATGTGATCA  | 2220 |
| CGGAGCCCGG  | CCTGACAGAG  | TTGTGGACCT | CGCCCGAAGT  | CGCCGAGGAC | CTCAAGGTAA  | 2280 |
| CTCTGAAAAA  | GAGGCAAGT   | CTTTTTTTTC | CCAACAAGAC  | AGTTGTGATC | TCTGGAGAGC  | 2340 |
| GCCATCGCTA  | TACGTGCGAG  | GTGCCGACGT | CGTCGCAAAAC | TTATAACATC | ACCAAGGGCT  | 2400 |
| TTAACTATAG  | CGCTCTGCCC  | GGGCACCTTG | GCGGATTTGG  | GATCAACGCG | CGTCTGGTAC  | 2460 |
| TGGGTGATAT  | CTTCGCATCA  | AAATGGTCGC | TATTCGCGAG  | GGACACCCCA | GAGTATCGGG  | 2520 |
| TGTTTTACCC  | AATGATTGTC  | ATGGCCGTCA | AGTTTTCCAT  | ATCCATTGGC | AACAACGAGT  | 2580 |
| CCGGCGTAGC  | GCTCTATGGA  | GTGGTGTCCG | AAGATTTTCG  | GGTCGTACCG | CTCCACAACA  | 2640 |
| GGTCCAAAGA  | GGCTAACGAG  | ACGGCGTCCC | ATCTTCTGTT  | CGGTCTCCCG | GATTCACTGC  | 2700 |
| CATCTCTGAA  | GGGCCATGCC  | ACCTATGATG | AACTCACGTT  | CGCCCGAAAC | GCAAAATATG  | 2760 |
| CGCTAGTGGC  | GATCCTGCCCT | AAAGATTCTT | ACCAGACACT  | CCTTACAGAG | AATTACACTC  | 2820 |
| GCATATTTCT  | GAACATGACG  | GAGTCGACGC | CCCTCGAGTT  | CACGCGGACG | ATCCAGACTA  | 2880 |
| GGATCGTATC  | AATCGAGGCC  | AGGCGCGCCT | GCGCAGCTCA  | AGAGGCGGGC | CCGGACATAT  | 2940 |
| TCTTGGTGT   | GTTTCAGATG  | TTGGTGGCAC | ACTTTCTTGT  | TGCGCGGGGC | ATTACCGAGC  | 3000 |
| ACCGATTTGT  | GGAGGTGGAC  | TGCGTGTGTC | GGCAGTATGC  | GGAAGTGTAT | TTTCTCCGCC  | 3060 |
| GCATCTCGCG  | TCTGTGCATG  | CCCACGTTCA | CCACTGTCCG  | GTATAACCAC | ACCACCCCTG  | 3120 |
| GCGCTGTGGC  | CGCCACACAA  | ATAGCTCGCG | TGTCCGCCAC  | GAAGTTGGCC | AGTTTGCCCC  | 3180 |
| GCTCTTCCCA  | GGAAACAGTG  | CTGGCCATGG | TCCAGCTTGG  | CGCCCGTGAT | GGCGCCGTCC  | 3240 |
| CTTCCTCCAT  | TCTGGAGGGC  | ATTGCTATGG | TGCTCGAACA  | TATGTATACC | GCCTACACTT  | 3300 |
| ATGTGTACAC  | ACTCGGCGAT  | ACTGAAAGAA | AATTAATGTT  | GGACATACAC | ACGGTCCCTCA | 3360 |
| CCGACAGCTG  | CCCCCCCCAA  | GACTCCGGAG | TATCAGAAAA  | GCTACTGAGA | ACATATTTGA  | 3420 |
| TGTTACATC   | AATGTGTACC  | AACATAGAGC | TGGGCGAAAT  | GATCGCCCGC | TTTTCCAAAC  | 3480 |
| CGGACAGCCT  | TAACATCTAT  | AGGGCATTCT | CCCCCTGCTT  | TCTAGGACTA | AGGTACGATT  | 3540 |
| TGCATCCAGC  | CAAGTTGCGC  | GCCGAGGCGC | CGCAGTCGTC  | CGCTCTGACG | CGGACTGCCC  | 3600 |
| TTGCCAGAGG  | AACATCGGGA  | TTCCGAGAAT | TGCTCCACGC  | GCTGCACCTC | GATAGCTTAA  | 3660 |
| ATTTAATTCC  | GGCGATTAAAC | TGTTCAAAGA | TTACAGCCGA  | CAAGATAATA | GCTACGGTAC  | 3720 |
| CCCTGCCTCA  | CGTCACGTAT  | ATCATCAGTT | CCGAAGCACT  | CTCGAACGCT | GTTGTCTACG  | 3780 |
| AGGTGTCCGA  | GATCTTCCTC  | AAGAGTGCCA | TGTTTATATC  | TGCTATCAAA | CCCGATTGCT  | 3840 |
| CCGGCTTTAA  | CTTTTCTCAG  | ATTGATAGGC | ACATTCCCAT  | AGTCTACAAC | ATCAGCACAC  | 3900 |
| CAAGAAGAGG  | TTGCCCCCTT  | TGTGACTCTG | TAATCATGAG  | CTACGATGAG | AGCGATGGCC  | 3960 |
| TGCAGTCTCT  | CATGTATGTC  | ACTAATGAAA | GGGTGCAGAC  | CAACCTCTTT | TTAGATAAGT  | 4020 |

|            |             |            |             |            |             |      |
|------------|-------------|------------|-------------|------------|-------------|------|
| CACCTTTCTT | CGATAATAAC  | AACCTACACA | TTCAATTATTT | GTGGCTGAGG | GACAAAGGGA  | 4080 |
| CCGTASTGGA | GATAAGGGGC  | ATGTATAGAA | GACGCGCAGC  | CAGTGCTTTG | TTTCTAATTC  | 4140 |
| TCTCTTTTAT | TGGGTTCTCG  | GCGGTATCT  | ACTTTCTTTA  | CAGACTGTTT | TCCATCCTTT  | 4200 |
| ATTAGACGGT | CAATAAAGCG  | TAGATTTTAA | AAAGGTTTTCC | TGTGCATTCT | TTTTGTATGG  | 4260 |
| GCATATACTT | GGCAAGAAAT  | CCGAGCACCT | CAGAAAGTGG  | ATTGCCGTCA | CATATCASTT  | 4320 |
| CGACCAACCC | TGCACCTAGC  | CATGCGGCGC | TTTGACGGTC  | TTTGGGGCTA | CACATCATAA  | 4380 |
| AGTACTTTTC | CATGGCTTCT  | ATAAGCACCT | TGGAACAATC  | TGGGGGTTGG | CGAATGGGTT  | 4440 |
| CCCTAAACCG | GAAATCCTCT  | ATGGTATTCA | GGCAGAAGAC  | CGCGTCCTCC | ACCCGACGTT  | 4500 |
| TGAGTCTTTC | TAGCAGAGCG  | CCGAAGAACT | CCCGCTCGTG  | TGTTTTCGCA | GGGGCAAGTT  | 4560 |
| CTGCGCCGTA | CAGCGATGAG  | AAACACGACA | CGATGTTTTT  | CAGCCCCATG | CTGCGCAGCA  | 4620 |
| ACACGTGCTT | CAGGAACAGG  | TGTTGTAGCC | GGTTCAGTTT  | TAGCTTGGGT | AGAAAAGTTA  | 4680 |
| TGAGTTGTTT | AGCACGCTCC  | ATGATGGTAA | CGGTGTTGAA  | GTCACAGACC | GGGCTTTCTC  | 4740 |
| CGAGTCTCGG | CCGCTTGAGT  | CCAATCATGT | AGAACATAGA  | CCCGGCTCCG | TTGTCTGTGT  | 4800 |
| TAAGTGACAC | GATATCCCGT  | TGCAAAACCT | GTGCGATGTT  | GTGTTTCAGT | ATAGATCTGG  | 4860 |
| TTTGACCGGC | ACGGGGTGTT  | ATGGGGTGAC | GCGGTAAAGG  | CGACTCTGGG | TCAAAACACCT | 4920 |
| TTATGCGGTT | GGCGGCTCCG  | TGATGACGA  | CACGCTTGTT  | CGCGGCGTGT | ATGGGGACGC  | 4980 |
| GACGGCATCC | CGCTGGCAGA  | TCTATAATCT | TAAAGTTGGT  | ATAAGACTGG | TGGCTGTCTA  | 5040 |
| TGGCCAGCCG | GCACTCCGGT  | AGTATCTGCG | TGTCTTCGAA  | TTGCTGGCCG | CGTACGACTG  | 5100 |
| GCTTGGAGTG | CAGGTAAACG  | CCAAGAGATG | CGGTCTCTTC  | GCTACGCCAC | AAGTGGCTTC  | 5160 |
| TTAACGCGTA | GGGCTGCGGT  | GAGAGCATGA | TCCGTAGCAA  | CGATAGTTCC | GGGTGCTTAG  | 5220 |
| CCGCGTAGAG | TGGCAGGGTA  | GACGAGTCCG | GAGTCCCAAA  | CTTTTCGAAC | AACAGTGGCA  | 5280 |
| TGCGGACTTC | AGGATTAGAG  | ACTCCCAACA | TGGCCGCCAC  | CGCCGGAGAG | GTCAAGACGT  | 5340 |
| GAAACACGCG | CTCGCTGTCT  | GACAGGCCTG | CCGCGCCCTC  | TACTAGACTA | GCCTTCACGT  | 5400 |
| CCGGAACCTG | TAAATAGCTT  | TAGACCAGCG | GACGGACGCA  | ACGTACSTGG | GGATCGGCTG  | 5460 |
| GCGGTGTCTG | CTCGTTGGAC  | GCGGCCGTTT | GGTGGCGCCA  | GTGCAGGCCT | AGTTTGCGAA  | 5520 |
| TGGCGTGACG | GACAAATTTGT | GGCTTTAGAG | CGGCGAACCG  | ATGACCCGTG | GTGGCGACGA  | 5580 |
| ACGAAATGAA | GTTTGCAATTG | CGGCCCCAAT | CGTCTAGCCT  | GCTCTTCTTG | TTTCCGGCAT  | 5640 |
| AGATTTTCCG | GATTAGGTTA  | CATTTTATAT | ATCCCACTAC  | TGCGCACTCG | TGTTTGCTTT  | 5700 |
| TASTGTGACT | GATTATCTTC  | TTTGAGAAGT | CAAACAGGCC  | CCGCGCGGCG | GCTCGCCTAA  | 5760 |
| TGCAAGCCAC | GTCAAGCCTG  | AGAAACGAAC | AGCATTCCAC  | CAGACACTCC | AGGAACCTTT  | 5820 |
| TGTGTAGCCT | CTGTATTTGG  | GAACGGTTTC | TGTGCTCAAG  | TAGGGAGAAT | ATTCTATTTT  | 5880 |
| TGTTTCCGTC | GATGCGCGCG  | TGCTGGTCCG | TGAGAATGGG  | CGCCAGCTCG | TGCGCAATCT  | 5940 |
| GTTCCACAAG | AGGCTGCCCC  | TACACTTTAG | AAATCGTGCC  | TGTCGCGGCC | TTAAACCAGG  | 6000 |
| ACACGTTTAG | CCCATCCTTG  | CTGGAGACCA | CAGATGGAAA  | GTTTGTGGTC | CAAAATACGT  | 6060 |



|             |             |            |            |            |             |      |
|-------------|-------------|------------|------------|------------|-------------|------|
| TTTTTGGCCC  | CATTCTCACC  | ATGTACTGGT | TTTCCAGTCC | GTGCAGGTCC | AACGTGGAGT  | 6120 |
| TCCAAATTTGC | TATCGATACA  | GGAAATATGT | GCCTGATTGG | CAGAAAGCAT | TTCCAGCGTAC | 6180 |
| CCATTGCGAA  | GAGAAAGTGC  | AGCATGTCCC | CACTGATGTT | GATGTTTATT | GCSTTGGCTT  | 6240 |
| GACACATGTT  | GTGGGAAAAA  | AACACGCTTA | TGGTAAAAGA | AGGTTCCCTT | ACGGASTACT  | 6300 |
| TTCTGTATAAC | AAAAATTGTTG | GTCAATCTGG | GGATGTTTAA | AATAGTCTTT | TCCAGGGTGT  | 6360 |
| TAGGAACGTG  | GCAGCTTATC  | TTAGTGTTAA | TCACCATGTT | GGTGTGAAT  | ATGGTGATCT  | 6420 |
| TGAAGTTTTTC | CAAACTGACC  | TGTTTTGTGG | GTTCCAGCAT | GTCTGACACT | GTAGAGCTGC  | 6480 |
| CCAGAGTCCC  | CGCGTCCGTG  | GCCGCGTATC | GTTGGAAGCA | CGCCTGCAAA | TTTCCCTTCA  | 6540 |
| TGGCTGCTCG  | CCGGTCTTTT  | GGCGCGTACC | GGATTCTTGA | AAGCGTCGCC | GCCAGGAGAC  | 6600 |
| GCGGTGTCTC  | GTGGGTGCCT  | AAAAAGTTTG | CGCAGGGGTG | CAGTCCGCTG | CACGAGTGGC  | 6660 |
| CGATGCAGTC  | TGCCACTGCC  | ATACACATGA | CGAGTCTGTA | GATGGCCGGT | GTGCCCGGAT  | 6720 |
| ACACTAGATA  | GTAGGTACAA  | TCTGGGGTAC | TGACGACCAC | CCTGTATGGC | TTTGGTCCGG  | 6780 |
| GGTCCCTTCCG | TTGGATTTTT  | ACGTGCAGAC | GGGACACGAG | CTGGTTTAGA | GCCAGCTGAA  | 6840 |
| AGCCCACCAG  | ATCCCGTCCG  | TTAACCTTGA | CGTCCTGGTG | CTTACTCTGT | TTCCAGAGGT  | 6900 |
| TCTTCAGCAC  | GGTGGGCAST  | CGCTCTACGT | TGTGAGCGAT | GGCACGGCCG | AGCGAGACCA  | 6960 |
| GCTCTCCGTG  | CCACCCCCAC  | GTGGCCATGA | AGCTGCTGAT | GTTAAACTTT | AAAAATGTA   | 7020 |
| GCTGTGCGTC  | TGGGGATGCG  | GGTGGCATT  | TTGAAAACGA | GAGATGCTTC | AGGCTCTCCA  | 7080 |
| GGAGTSCAAA  | ATAATTTTGA  | TAGATTGTGG | GTTGTAGACT | ATGGGSCAAC | ACCGCCAGAA  | 7140 |
| ACGCATGAAA  | ACACTGTTCC  | AACTCCAGTA | ACTCCAGGTA | CCTGCACACT | ATCCTGAACA  | 7200 |
| TGGCTTTGTA  | ACATATGGTG  | CACGTTAGTA | GCGCGGGAAG | ATACAGCGAG | CGTAGCTCCC  | 7260 |
| TGAATTCGCA  | GGGTTTATCA  | CAATCATCGG | TAAGTTCCCA | TGATCCCAAC | GCAGGTAGGT  | 7320 |
| AGTTGTGCGT  | GTCTATCTGT  | CCGCGCGTAA | ACACTCCACC | ACCGTCAATT | ATTAAACCTT  | 7380 |
| CGCCGCTGTA  | CCGTGACCCC  | ACTTTTCCCA | AAAGAGTCCC | TTCTTGATGT | ATAAAGGGT   | 7440 |
| GGAGGCGCTT  | CCCCAGGAGT  | AGTCTGCGTA | TCGCTCTGCA | GSCGAAAAAG | GTGGGCTCGG  | 7500 |
| GCTGCATCAT  | CTTATCAAGA  | CCTTCTAAGG | TCAGCTCTGC | CTGCAGGTGC | GAGTTGGTGG  | 7560 |
| CCAGACAGCA  | GAATATTTCC  | AGCTGTGATT | CCCAAGTCGC | TTGATAACAC | GTGGTCTGCG  | 7620 |
| GACTCGTCGT  | CAGGGAGGCG  | CTCGGTGGCA | GTAGTAGGGG | GCCCTCGAGC | GCTGCCATGG  | 7680 |
| AGGCGACCTT  | GGAGCAACGA  | CCTTTCCCGT | ACCTCGCCAC | GGAGGCCAAC | CTCCTAACGC  | 7740 |
| AGATTAAGGA  | GTGGGCTGCC  | GACGGACTCT | TCAAGAGCTT | TCAGCTATTG | CTCGGCAAGG  | 7800 |
| ACGCCAGAGA  | AGGCAGTGTC  | CGTTTCGAAG | CGCTACTGGG | CGTATATACC | AATGTGSTGG  | 7860 |
| AGTTTGTTAA  | GTTTCTGGAG  | ACCGCCCTCG | CCGCCGCTTG | CGTCAATACC | GAGTTCAAGG  | 7920 |
| ACCTGCGGAG  | AATGATAGAT  | GGAAAAATAC | AGTTTAAAT  | TTCAATGCCC | ACTATTGCCC  | 7980 |
| ACGGAGACGG  | GAGGAGGCC   | AACAAGCAGA | GACAGTATAT | CGTCATGAAG | GCTTGCAATA  | 8040 |
| AGCACCACAT  | CGGTGCGGAG  | ATTGAGCTTG | CGGCCGCAGA | CATCGAGCTT | CTCTTCGCGG  | 8100 |

|            |             |             |            |             |             |       |
|------------|-------------|-------------|------------|-------------|-------------|-------|
| AGAAAGAGAC | GCCCTTGGAC  | TTACACAGAGT | ACGGCGGTGG | CATCAAGAGG  | ATTACGTGGG  | 8160  |
| CTTTGCAATT | TGCTATGGAC  | GCCCTAGAAC  | GGGGGTTAGT | GGACACGGTT  | CTCGCAGTTA  | 8220  |
| AACCTCGGCA | CGCTCCACCC  | GTCTTTATTT  | TAAAGACGGT | GGGCGATCCC  | GTCTACTCTG  | 8280  |
| AGAGGGGCGT | CAAAAAGGCC  | GTCAAGTCTG  | ACATGGTATC | CATGTTCAAG  | GCACACCTCA  | 8340  |
| TAGAACATTC | ATTTTTTCTA  | GATAAGGCGG  | AGCTCATGAC | AAGGGGGAAG  | CAGTATGTCC  | 8400  |
| TAACCATGCT | CTCCGACATG  | CTGGCTGCGG  | TGTGCGAGGA | TACCGTCTTT  | AAGGGTGTCA  | 8460  |
| GCACGTACAC | CACGGGCTCT  | GGGCAGCAGG  | TGGCCGGCGT | CCTGGAGACG  | ACGGACAGCG  | 8520  |
| TCATGAGACG | GCTGATGAAC  | CTGCTGGGGC  | AAGTGGAAAG | TGCCATGTCC  | GGGCCCGCGG  | 8580  |
| CCTACGCCAG | CTACGTTGTC  | AGGGGTGCCA  | ACCTCGTCAC | CGCCGTTAGC  | TACGGAAGGE  | 8640  |
| CGATGAGAAA | CTTTGAACAG  | TTTATGGCAC  | GCATAGTGGG | CCATCCCAAC  | GCTCTGCCGT  | 8700  |
| CTGTGGAAGG | TGACAAGGCC  | GCTCTGGCGG  | ACGGACACGA | CGAGATTCCG  | AGAACCCGCA  | 8760  |
| TGCGCGCCTC | TCTCGTCAAG  | ATAGGGGATA  | AGTTTGTGGC | CATTGAAAST  | TTGCAGCGCA  | 8820  |
| TGTACAACGA | GACTCAGTTT  | CCCTGCCCAC  | TGAACCGGCG | CATCCAGTAC  | ACCTATTTCT  | 8880  |
| TCCCTGTTGG | CCTTCACCTT  | CCCGTGCCCC  | GCTACTCGAC | ATCCGTCTCA  | GTGAGGGGCG  | 8940  |
| TAGAATCCCC | GGCCATCCAG  | TCGACCGAGA  | CGTGGGTGGT | TAATAAAAC   | AACGTGCCCTC | 9000  |
| TTTGCTTCGG | TTACCAAAAC  | GCCCTCAAAA  | GCATATGCCA | CCCTCGAATG  | CACAAACCCA  | 9060  |
| CCCAGTCAGC | CCAGGCACTA  | AACCAAGCTT  | TTCCCGATCC | CGACGGGGGA  | CATGGGTACG  | 9120  |
| GTCTCAGGTA | TGAGCAGACG  | CCAAACATGA  | ACCTATTCAG | AACGTTCCAC  | CAGTATTACA  | 9180  |
| TGGGGAAAAA | CGTGGCATTG  | GTTCCCGATG  | TGGCCCCAAA | AGCGCTCGTA  | ACCACGGAGG  | 9240  |
| ATCTACTGCA | CCCAACCTCT  | CACCGTCTCC  | TCAGATTGGA | GGTCCACCCC  | TTCTTTGATT  | 9300  |
| TTTTTGTTGA | CCCTGTCTCT  | GGAGCGAGAG  | GATCGTACCG | CGCCACCCAC  | AGAACAATGG  | 9360  |
| TTGGAATAT  | ACCACAAACG  | CTCGCTCCAA  | GGGAGTTTCA | GGAAAGTAGA  | GGGGGCGCAGT | 9420  |
| TCGACGCTGT | GACGAATATG  | ACACACGTCA  | TAGACCAGCT | AACATTTGAC  | GTCTACAGG   | 9480  |
| AGACGGCATT | TGACCCCGCG  | TATCCCTGTG  | TCTGCTATGT | AATCGAAGCA  | ATGATTACCG  | 9540  |
| GACAGGAAGA | AAAATTCTGT  | ATGAACATGC  | CCCTCATTCG | CCTGGTCAAT  | CAACCTACT   | 9600  |
| GGGTCAACTC | GGGAAAACCTG | GCGTTTGTGA  | ACAGTTATCA | CATGGTTAGA  | TTCATCTGTA  | 9660  |
| CGCATATGGG | GAATGGAAGC  | ATCCCTAAGG  | AGGCGCACGG | CCACTACCGG  | AAAATCTTAG  | 9720  |
| GCGAGCTCAT | CGCCCTTGAG  | CAGGCGCTTC  | TCAAGCTCGC | GGGACACGAG  | ACGGTGGGTC  | 9780  |
| GGACGCCGAT | CACACATCTG  | GTTTCGGGTC  | TCCTCGACCC | GCATCTGCTG  | CCTCCCTTTG  | 9840  |
| CCTACCACGA | TGTCTTTACG  | GATCTTATGC  | AGAAGTCATC | CAGACAACCC  | ATAATCAAGA  | 9900  |
| TCGGGGATCA | AAACTACGAC  | AACCCCTCAA  | ATAGGGCGAC | ATTCAATCAAC | CTCAGGGGTC  | 9960  |
| GCATGGAGGA | CCTAGTCAAT  | AACCTTGTTA  | ACATTTACCA | GACAAGGGTC  | AATGAGGACC  | 10020 |
| ATGACGAGAG | ACACGTCTCT  | GACGTGGCGC  | CCCTGGACGA | GAATGACTAC  | AACCCGGTCC  | 10080 |
| TCGAGAAGCT | ATTCTACTAT  | GTTTTAATGC  | CGGTGTGCAG | TAACGGCCAC  | ATGTGCGGTA  | 10140 |

|            |            |            |             |            |            |       |
|------------|------------|------------|-------------|------------|------------|-------|
| TGSGGGTGGG | STATCAAAAC | GTGSCCCTGA | CGCTGACTTA  | CAACGGCCCC | GTCTTTGGGG | 10200 |
| ACGTGCTGAA | CGCACAGGAT | GATATTCTAC | TGCACCTGGA  | GAACGGAAAC | TTGAAGGACA | 10260 |
| TTCTGCAGGC | AGGCGACATA | CGCCCCGACG | TGGACATGAT  | CAGGCTGCTG | TGCACCTCCT | 10320 |
| TTCTGACGTG | CCCTTTCTGT | ACCCAGGCGG | CTCGCGTGAT  | CACAAAGCGG | GACCCGGCCC | 10380 |
| AGAGTTTTTG | CACGCACGAA | TACGGGAAGG | ATGTGGCGCA  | GACCGTGCTT | GTTAATGSGT | 10440 |
| TTGGTGCGTT | CGCGGTGGCG | GACCGCTCTC | GCGAGGCGGC  | GGAGACTATG | TTTTATCCGG | 10500 |
| TACCCCTTAA | CAAGCTCTAC | GCTGACCCGT | TGGTGGCTGC  | CACACTGCAT | CCGCTCCTGG | 10560 |
| CAAACTATGT | CACCAGGCTC | CCCAACCAGA | GAAACGCGGT  | GGTCTTTAAC | GTGCCATCCA | 10620 |
| ATCTCATGGC | AGAATATGAG | GAATGGCACA | AGTCGCCCGT  | CGCGGCGTAT | GCCGCGTCTT | 10680 |
| GTCAGGCCAC | CCCGGGCCGC | ATTAGCGCCA | TGGTGAGCAT  | GCACCAAAAA | CTATCTGCCC | 10740 |
| CCAGTTTCAT | TTGCCAGGCA | AAACACCGCA | TGCACCCCTG  | TTTTGCCATG | ACAGTCTGCA | 10800 |
| GGACGGACGA | GTTTCTAGCA | GAGCACATCC | TATACTGCTC  | CAGGGCGCTC | ACATCCATGT | 10860 |
| TTGTGGGCTT | GCCTTCGGTG | GTACGGCGCG | AGGTACGTTT  | GGACGCGGTG | ACTTTTGAAA | 10920 |
| TTACCCACGA | GATCGCTTCC | CTGCACACCG | CACCTTGGCTA | CTCATCAGTC | ATCGCCCCGG | 10980 |
| CCCACGTGGC | CGCCATAACT | ACAGACATGG | GAGTACATTG  | TCAGGACCTC | TTTATGATTT | 11040 |
| TCCCAGGGGA | CGCGTATCAG | GACCGCCAGC | TGCATGACTA  | TATCAAAATG | AAAGCGGGCG | 11100 |
| TGCAAAACCG | CTCACCGGGA | AACAGAATGG | ATCACGTGGG  | ATACACTGCT | GGGGTTCCTC | 11160 |
| GCTGCGAGAA | CCTGCCCCGT | TTGAGTCATG | GTCAGCTGGC  | AACCTGCGAG | ATAATTCCCA | 11220 |
| CGCCGGTCA  | ATCTGACGTT | GCCTATTTCC | AGACCCCCAG  | CAACCCCCGG | GGCGGTGCGG | 11280 |
| CCTGCGTGGT | GTCGTGTGAT | GCTTACAGTA | ACGAAAGCGC  | AGAGCGTTTG | CTCTACGACC | 11340 |
| ATTCAATACC | AGACCCCCGG | TACGAATGCC | GGTCCACCAA  | CAACCCGTGG | GCTTCGCAGC | 11400 |
| GTGGCTCCCT | CGGCGACGTG | CTATACAATA | TCACCTTTCC  | CCAGACTGCG | CTGCCGGGCA | 11460 |
| TGTACAGTCC | TTGTGGGCG  | TTCTTCCACA | AGGAAGACAT  | TATGCGGTAC | AATAGGGGGT | 11520 |
| TGTACACTTT | GGTTAATGAG | TATTCTGCCA | GGCTTGCTGG  | GGCCCCGGCC | ACCAGCACTA | 11580 |
| CAGACCTCCA | GTACGTCTGT | GTCAACGGTA | CAGACGTGTT  | TTTGGACCAG | CCTTGCCATA | 11640 |
| TGCTGCAGGA | GGCCTATCCC | ACGCTCGCCG | CCAGCCACAG  | AGTTATGCTT | GACGAGTACA | 11700 |
| TGTCAAACAA | GCAGACACAC | GCCCCAGTAC | ACATGGGCCA  | GTATCTCATT | GAAGAGGTGG | 11760 |
| CGCCGATGAA | GAGACTATTA | AAGCTCGGAA | ACAAGGTGGT  | GTATTAGCTA | ACCTTTCTAG | 11820 |
| CGTTGGCTAG | TCATGGCACT | CGACAAGAGT | ATAGTGGTTA  | ACTTCACCTC | CAGACTCTTC | 11880 |
| GCTGATGAAC | TGGCCGCCCT | TCAGTCAAAA | ATAGGGAGCG  | TACTGCCGCT | CGGAGATTGC | 11940 |
| CACCGTTTAC | AAAAATATAC | GGCATTGGGC | CTGGGGTGCG  | TATGCTCACC | TGAGACATCT | 12000 |
| CCGGACTACA | TCCAAATTAT | GCAGTATCTA | TCCAAGTGCA  | CACTCGCTGT | CCTGGAGGAG | 12060 |
| GTTCCGCCCG | ACAGCCTGCG | CCTAACGCGG | ATGGATCCCT  | CTGACAACCT | TCAGATAAAA | 12120 |
| AACGTATATG | CCCCCTTTTT | TCAGTGGGAC | AGCAACACCC  | AGCTAGCAGT | GCTACCCCCA | 12180 |

|            |            |            |            |            |            |       |
|------------|------------|------------|------------|------------|------------|-------|
| TTTTTTAGCC | GAAAGGATTC | CACCATTTGT | CTCGAATCCA | ACGGATTGGA | CCTCGTGTTC | 12240 |
| CCCATGGTGG | TGCCCAGCA  | ACTGGGGCAC | GCTATTCTGC | AGCAGCTGTC | GGTGTACCAT | 12300 |
| ATCTACTCCA | AAATATCGGC | CGGGGCCCCG | GATGATGTAA | ATATGGCGGA | ACTTGATCTA | 12360 |
| TATACCACCA | ATGTGTCAAT | TATGGGGGCG | ACATATCGTC | TGGACGTAGA | CAACACGAT  | 12420 |
| CCACGTACTG | CCCTGCGAGT | GTTTGACGAT | CTGTCCATGT | ACCTTTGTAT | CCTATCAGCC | 12480 |
| TTGGTTCCCA | GGGGGTGTCT | CCGTCTGCTC | ACGGGCGCTG | TGGGGCACGA | CAGGCATCCT | 12540 |
| CTGACAGAGG | TGTTTGAGGG | GTTGGTGCCA | GATGAGGTGA | CCAGGATAGA | TCTCGACCAG | 12600 |
| TTGAGCGTCC | CAGATGACAT | CACCAGGATG | CGCGTCAATG | TCTCCTATCT | TCAGATCTCT | 12660 |
| AGTTCTATAT | TTAATCTTGG | CCCCAGACTG | CACGTGTATG | CCTACTCGGC | AGAGACTTTG | 12720 |
| GCGGCCTCCT | GTGGGTATTC | CCCACGCTAA | CGATTTGAAG | CGGGGGGGGG | GTATGGCGTC | 12780 |
| ATCTGATATT | CTGTGGTTTG | CAAGGACGGA | TGACGGCTCC | GTCTGTGAAG | TCTCCCTGCG | 12840 |
| TGGAGGTAGG | AAAAAACTA  | CGTCTACCT  | GCCGGACACT | GAACCTGGG  | TGCTAGAGAC | 12900 |
| CGACGCCATC | AAAGACGCTT | TGCTCAGCGA | CGGGATCGTG | GATATGGCTC | GAAAGCTTCA | 12960 |
| TCGTGGTGCC | CTGCCCTCAA | ATTCTCAGAA | CGGCTTGAGG | ATGGTGCTTT | TTTGTATTGG | 13020 |
| TTACTTGCAA | AATTGTGTGT | ACCTAGCCCT | GTTTCTGTGC | CCCTTAATC  | CTTACTTGGT | 13080 |
| AACTCCCTCA | AGCATTGAGT | TTGCCGAGCC | CGTTGTGGCA | CCTGAGGTGC | TCTTCCACAA | 13140 |
| CCCGGCTGAG | ATGTCTCGCG | GTTGGGATGA | CGCGATTTTC | TGTAACTGC  | CCTATACCGT | 13200 |
| GCCTATAATC | AACACCACGT | TTGGACGCAT | TTACCCGAAC | TCTACACGGG | AGCCGGACGG | 13260 |
| CAGGCCTACG | GATTACTCCA | TGGCCCTTAG | AAGGGCTTTT | GCAGTTATGG | TTAACACGTC | 13320 |
| ATGTGCAGGA | GTGACATTGT | GCCGCGGAGA | AACTCAGACC | GCATCCCGTA | ACCACACTGA | 13380 |
| GTGGGAAAAT | CTGCTGGCTA | TGTTTTCTGT | GATTATCTAT | GCCTTAGATC | ACAACTGTCA | 13440 |
| CCCGGAAGCA | CTGTCTATCG | CGAGCGGCAT | CTTTGACGAG | CGTGAATATG | GATTATTCTT | 13500 |
| CTCTCAGCCC | CGGAGCGTGC | CCTCGCCTAC | CCCTTGCGAC | GTGTCTGGG  | AAGATATCTA | 13560 |
| CAACGGGACT | TACCTAGCTC | GGCCTGGAAA | CTGTGACCCC | TGGCCCAATC | TATCCACCTC | 13620 |
| TCCCTTGATT | CTAAATTTTA | AATAAAGGTG | TGTCACTGGT | TACACCACGA | TTAAAAACCA | 13680 |
| CTCACTGAGA | TGTCTTTTTA | ACCGCTAAGG | GATTATACCG | GGATTTAAAA | CCGCCCACTG | 13740 |
| ATTTTTTTAC | GCTAAGAGTT | GGGTGCTTGG | GGGGTTTTGC | ATTGCTCTGT | TGTAACTAT  | 13800 |
| ATATAAGTTA | AACCAAAATT | CGCAGGGAGA | CAAGGTGACG | GTGGTGAGAA | CTCAGTTGAG | 13860 |
| AGTCAGAGAA | TACAGTGCTA | ATCAGGGTAG | ATGAGCATGA | CTTCCCGCTC | TCCAGTCAAC | 13920 |
| GGAGGAATGG | TGGACGGGTC | CGTCTGGGTG | CGAATGGCCA | CCAAGCCTCC | CGTGATTGGT | 13980 |
| CTTATAACAG | TGCTCTTCCT | CCTAGTCATA | GGCGCCTGCG | TCTACTGCTG | CATTGCGCTG | 14040 |
| TTCTTGGCGG | CTCGACTGTG | GCGCGCCACC | CCACTAGGCA | GGGCCACCGT | GCGGTATCAG | 14100 |
| GTCTTTCGCA | CCCTGGGACC | GCAGGCCGGG | TCACATGCAC | CGCCGACGGT | GGGCATAGCT | 14160 |
| ACCCAGGAGC | CCTACCGTAC | AATATACATG | CCAGATTAGA | ACGGGGTGTG | TGCTATAATG | 14220 |

|             |            |            |            |            |            |       |
|-------------|------------|------------|------------|------------|------------|-------|
| GATGGCTATG  | GGGGGGCTGT | AGATAATTGA | GCGCTGTGCT | TTTATTGTGG | GGATATGGGC | 14280 |
| TTGTACATGT  | GTCTATCATC | GGTAGCCATA | AAATGGGCCA | TGACAACTGC | CACAASTAAG | 14340 |
| TGCTCCGACA  | TGTGCTTTTG | CTTGGCGCTG | TATGACTGCC | CTCCATCCCT | AAGCGGGAGC | 14400 |
| CACCTTGATCG | CGCGGACCTG | TTCTACCAGG | TAGGTACCGG | GGTCAAATGA | TATTTTGATG | 14460 |
| GTGTTGGACA  | CCACCGTCTG | GCTGGCGCTC | AGGGTGCCGG | AGTTGAGAGC | GTAGATGAAT | 14520 |
| GTCTCAAACG  | CGGAGGATTT | CTCGCCTCCC | AACATGTAAA | TTGGCCACTG | CAGGGCGCTG | 14580 |
| CTCTTGTCAG  | TATAGTGTAG | AAAATGTATG | GGGAGCGGGC | ATATTTCTGT | AAGGACGGTT | 14640 |
| GCAATGGGCA  | CCCCAGAATC | TTGGCTGCTG | TTGCCTTCGA | CCGCCGCGTT | CACGCGCTCA | 14700 |
| ATTGTGGGCT  | GGAGCACAGC | GATCGCCTTA | ATCATCGTGC | ATGCGCAGGA | CGCTATCTCG | 14760 |
| TAAGCAGCTG  | CGCCASTGAG | GTCGCGCAGG | AAGAAATGCT | CCATGCCCAA | TATGAGGCTT | 14820 |
| CTGGTGGGAG  | TCTGAGTACT | CGTGACAACG | GCGCCCACGC | CAGTACCGGA | CGCCTCCGTG | 14880 |
| TTGTTCTGTAT | ACGCGGGGTG | GATGTAAACA | AACAGCTGTT | TTCCAAGGCA | CTTCTGAACC | 14940 |
| TGCTGGGCGG  | TGGTGTCTAC | CCGACACATG | TCAAACTGTG | TCAGCGCTGC | GTACCCACCC | 15000 |
| ACGCGSTAAA  | GCGTAGCATT | TGACGACGCT | GCTCCCTCGC | CCATTAGTTC | GGTGTGGAAT | 15060 |
| GCCCCCTCCA  | TAAAGAGGTT | GGTGGTGGTT | TTGATGGATT | CGTCGATGGT | GATGTACGTC | 15120 |
| GGAATGTGCA  | GTCTGTAACA | AGGACAGGAC | ACTAGTGCGT | CTTGCAGGTG | GAAATCTTGG | 15180 |
| CGGTGGTCCG  | CACACACGTA | ACTGACCACA | TTGAGCATCT | TTTCTGGGGC | GTTCTTGAGG | 15240 |
| TTAAGCAGGA  | AACTCGTGGA | GCGGTCTGAC | GAGTTCACGG | ATGATATAAA | TATAAGCTTG | 15300 |
| GCGTCTTTCT  | GAAGCATGAA | ACCCAGAATA | GCCGGCAGTG | CATCCTTTTT | AATAAAATTC | 15360 |
| GCTCTGTCTA  | CGTAGAGCAG | GTTAAAGGTC | TGTCCCGGAA | TGCTCTGCAG | ACACGGAAAG | 15420 |
| ACACAAAAGA  | GGGGCTCATA | AGCGGCTAAC | AGTAAAGGAG | AGGAGGCGAA | CAGTGCCTGG | 15480 |
| CTCTGTCTCT  | TGGGAATAAA | AGGGGGCGTG | TGTGCCGATC | GTATGGGTGA | GCCAGTGGAT | 15540 |
| CCTGGACATG  | TGGTGAATGA | GAAAGATTTT | GAGGAGTGTG | AACAATTTTT | CAGTCAACCC | 15600 |
| CTTAGGGAGC  | AAGTGGTCCG | GGGGGTCAGG | GCACTCGACG | GCCTCGGTCT | CGCTGACTCT | 15660 |
| CTATGTCACA  | AAACAGAAAG | ACTCTGCCCT | CTGATGGACC | TGGTGGGCAC | GGAGTGCTTT | 15720 |
| GCGAGGGTGT  | GCCGCCTAGA | CACCGGTGCG | AAATGAAGAG | TGTGGCGAGT | CCCTTATGTC | 15780 |
| AGTTCCACGG  | CGTGTTTTGC | CTGTACCACT | GTGCCCAGTG | CCTGGCATAC | CACGTGTGTG | 15840 |
| ATGGGGGGCG  | CGAATGCGTT | CTCCTGCATA | CGCCGGAGAG | CGTCATCTGC | GAACTAACGG | 15900 |
| GTAAGTGCAT  | GCTCGGCAAC | ATTCAAGAGG | GCCAGTTTTT | AGGGCCGGTA | CCGTATCGGA | 15960 |
| CTTTGGATAA  | CCAGGTTGAC | AGGGACGCAT | ATCACGGGAT | GCTAGCGTGT | CTGAAACGGG | 16020 |
| ACATTGTGCG  | GTATTTGCAG | ACATGGCCGG | ACACCACCGT | AATCGTGCAG | GAAATAGCCC | 16080 |
| TGGGGGACGG  | CGTCACCGAC | ACCATCTCGG | CCATTATAGA | TGAAACATTC | GGTGAGTGTC | 16140 |
| TTCCCGTACT  | GGGGGAGGCC | CAAGGCGGGT | ACGCCATGGT | CTGTAGCATG | TATCTGCACG | 16200 |
| TTATCGTCTC  | CATCTATTCT | ACAAAAACGG | TGTACAACAG | TATGCTATTT | AAATGCACAA | 16260 |

|             |            |             |            |            |            |       |
|-------------|------------|-------------|------------|------------|------------|-------|
| AGAATAAAAA  | GTACGACTGC | ATTGCCAAGC  | GGGTGCGGAC | AAAATGGATG | CGCATGCTAT | 16320 |
| CAACGAAAGA  | TACGTAGGTC | CTCGCTGCCA  | CCGTTTGGCC | CACGTGGTGC | TGCGTAGGAC | 16380 |
| CTTTCTGCTG  | CATCAGGSCA | TACCCCTGGA  | GCCCGAGATC | ATCTTTTCCA | CCTACACCCG | 16440 |
| GTTCAGCCCG  | TGCGCAGGGT | CATCCCGCCG  | GTTGGTGGTG | TGTGGGAAAC | GTGTCTGTCC | 16500 |
| AGGGGAGGAA  | AACCAACTTG | CGTCTTCACC  | TTCTGGCTTG | GCGCTTAGCC | TGCGTCTGTT | 16560 |
| TTCCACAGAT  | GGGAACTTTC | ATCCATTTGA  | CATCTCGGTA | CTGCGCATTT | CCTGCCCTGG | 16620 |
| TTCTAATCTT  | AGTCTTACTG | TCAGATTTCT  | CTATCTATCT | CTGGTGGTGG | CTATGGGGGC | 16680 |
| GGGACGGAAT  | AATGCGCGGA | GTCCGACCCT  | TGACGGGGTA | TGCGCGCCAG | AGGGCGCCCT | 16740 |
| AGCCCAACCT  | TTGGAGGAAC | TGCAGAGGCT  | GGCGCGTGCT | ACGCCGGACC | CGGCACTCAC | 16800 |
| CCGTGGACCG  | TTGCAGGTCC | TGACCGGCCT  | TCTCCGCGCA | GGGTGAGACG | GAGACCGCGC | 16860 |
| CACTCACCAC  | ATGGCGCTCG | AGGCTCCGGG  | AACCGTGCGT | GGAGAAAGCC | TAGACCCGCC | 16920 |
| TGTTTCACAG  | AAGGGGCCAG | CGCGCACACG  | CCACAGGCCA | CCCCCGTGC  | GACTGAGCTT | 16980 |
| CAACCCCGTC  | AATGCTGATG | TACCCGCTAC  | CTGGCGAGAC | GCCACTAAGC | TGTACTCGGG | 17040 |
| TGCTCCCTAC  | TATGTGTGTG | TTTACGAACG  | CGGTGGCCCT | CAGGAAGACG | ACTGGGTGCC | 17100 |
| GATACCACTG  | AGCTTCCCAG | AAGAGCCCGT  | GCCCCCGCCA | CCGGGCTTAG | TGTTCTATGA | 17160 |
| CGACTTGCTC  | ATTAACACGA | AGCAGTGCGA  | CTTTGTGGAC | ACGCTAGAGG | CCGCCTGTGG | 17220 |
| CACGCAAGGC  | TACACGTTGA | GACAGCGCGT  | GCCTGTGCGC | ATTCTCTCGG | ACGCGGAAAT | 17280 |
| CGCAGACGCA  | GTAAATCGC  | ACTTTTTAGA  | GGCGTGCTTA | GTGTTACGGG | GGCTGGCTTC | 17340 |
| GGAGGCTAGT  | GCCTGGATAA | GAGCTGCCAC  | GTCCCCGCCC | CTTGGCCGCC | ACGCCTGCTG | 17400 |
| GATGGACGTG  | TTAGGATTAT | GGGAAAGCCG  | CCCCCACACT | CTAGGTTTGG | AGTTACGCGG | 17460 |
| CGTAAACTGT  | GGCGGCACGG | ACGGTGACTG  | GTTAGAGATT | TAAACACAGC | CCGATGTGCA | 17520 |
| AAAGACAGTC  | ACCGGGAGTC | TTGTGGCATG  | CGTGATCGTC | ACACCCGCAT | TGGAAACCTG | 17580 |
| GCTTGCTGTTA | CCTGGGGGTT | TTGCTATTAA  | AGGCCGCTAT | AGGGCGTCCA | AGGAGGATCT | 17640 |
| GGTGTTCATT  | CGAGGCCGCT | ATGCTAGCC   | GGAGGCGCAA | ACTTCGGAAT | TTCTAAACA  | 17700 |
| AGGAATGCAT  | ATGGACTGTT | AACCCAATGT  | CAGGGGACCA | TATCAAGGTC | TTTAACGCCT | 17760 |
| GCACCTCTAT  | CTCGCCGGTG | TATGACCCTG  | AGCTGGTAAC | CAGCTACGCA | CTGAGCCTGC | 17820 |
| CTGCTTACAA  | TGTGTCTGTG | GCTATCTTGC  | TGCATAAAGT | CATGGGACCG | TGTGTGGCTG | 17880 |
| TGGGAATTAA  | CGGAGAAATG | ATCATGTACG  | TCGTAAGCCA | GTGTGTTTCT | GTGCGGCCCG | 17940 |
| TCCCGGGGCG  | CGATGGTATG | GCCTCTCATCT | ACTTTGGACA | GTTTCTGGAG | GAAGCATCCG | 18000 |
| GACTGAGATT  | TCCCTACATT | GCTCCGCCGC  | CGTCGCGCGA | ACACGTACCT | GACCTGACCA | 18060 |
| GACAAGAATT  | AGTTCATACC | TCCCAGGTGG  | TGCGCCGCGG | CGACCTGACC | AATTGCACTA | 18120 |
| TGGGTCTCGA  | ATTGAGGAAT | GTGAACCCCT  | TTGTTTGGCT | CGGGGGCGGA | TGGGTGTGGC | 18180 |
| TGCTGTTCTT  | GGGCGTGGAC | TACATGGCGT  | TCTGTCCGGG | TGTGACCGGA | ATGCCGTCTG | 18240 |
| TGGCAAGAGT  | GGCCGCCCTG | CTTACCAGGT  | GCGACCACCC | AGACTGTGTC | CACTGCCATG | 18300 |

|            |            |             |             |            |            |       |
|------------|------------|-------------|-------------|------------|------------|-------|
| GACTCCGTGG | ACACGTTAAT | GTATTTCTGTG | GGTACTGTTC  | TGCGCAGTCG | CCGGGTCTAT | 18360 |
| CTAACATCTG | TCCCTGTATC | AAATCATGTG  | GGACCGGGAA  | TGGAGTGACT | AGGGTCACTG | 18420 |
| GAAACAGAAA | TTTTCTGGGT | CTTCTGTTCG  | ATCCCATTTG  | CCAGAGCAGG | GTAACAGCTC | 18480 |
| TGAAGATAAC | TAGCCACCCA | ACCCCCACGC  | ACCTCGAGAA  | TGTGCTAACA | GGAGTGCTCG | 18540 |
| ACGACGGCAC | CTTGGTGCCG | TCCGTCCAAG  | GCACCCCTGGG | TCTCTTTACG | AATGTCTGAC | 18600 |
| TACTTCAGCC | GCTTGCTGAT | ATATGAGTGT  | AAAAAACTTA  | AGGCCCTGGG | CTTACGTTCT | 18660 |
| TATTGAAGCA | TGTTGCGCAC | ATCAGCGAGC  | TGGACCGTCC  | TCCGGGTCCG | GTGTAGATTA | 18720 |
| TGGTTCCGTT | CTCCTTCTTG | ATGTTTAAAT  | TTTTGGGGGG  | GAACCAACGA | CAAAGCGTCT | 18780 |
| TTATGATTTG | CGCGAACACG | GAGTTGGCTA  | CGTGCTTTTG  | GTGGGCTACG | TACCCAAATG | 18840 |
| TAATGTTCTC | TACGGATGCC | AGTAGCATGC  | TGATGATCGC  | CACCACTATC | CATGTCTTTG | 18900 |
| CGTGTCTCCT | TGGTATTAGG | AATACGTTTG  | CCTTTTGCTT  | AAACGTCTGT | AAAACACTGT | 18960 |
| TTGGAGTTTC | AAATAAAACG | AAGTACTGCT  | TAAACAATCC  | AAACAACTGG | TGCGTCTTTT | 19020 |
| GTGGGGCCTT | GATTGAAACC | AAAAAGAAAA  | AAGTGTGCAT  | TACTAGCTGC | TGTTGGAAGG | 19080 |
| GCTCCAGCCA | GTGCACCCCG | GGAACGTAAC  | AGCCGTTTCAG | AAAGGACGAA | AGGTTAACCA | 19140 |
| GAAAAGCCTG | AAGTTCCCGG | TAGACAGAGC  | AGGCGTGCAG  | GGAGTCGTGT | GTTTTCTGCG | 19200 |
| CCGCTCTGTA | CTCGACCACT | TGATCGGGCG  | TGGAGACGTG  | CGCGTCTCTG | CGCACACACC | 19260 |
| GCACTCTGCA | GTATGTTGAT | AGGGACTCCA  | ATAGGCGCGG  | CTTTGCGGGG | ACGTTGTCTT | 19320 |
| CGGACGGTCT | GGGGGTTCCT | ACGTCCGGAT  | TTGCTGACGT  | GGGCGTGGCG | GGATGGTGCC | 19380 |
| GTGTGCASTA | TGTTTCCAGG | ACCGAACTGT  | ATGAGTTTAT  | TCTGTGCACC | ACGCCAATAA | 19440 |
| AAGGGTGCGC | CATCCGTGCC | GTTTTGGGAC  | AGTGTCTCGT  | GAATGTCCGG | GCACTCAGTT | 19500 |
| CCCACCTCTC | TCCGGCGTCT | TTGGCGGTCT  | CCTGCAGGTT  | GGCGGCAAGG | CGCTCCCTGT | 19560 |
| GACGCTCTAG | CAGCATGTTT | GCTTTGAGCT  | CGCTCCTGTC  | CGAGGGTGAC | CCGGAGGTGA | 19620 |
| CCAGTAGGTA | CGTCAAGGGC | GTACAACTTG  | CCCTGGACCT  | TAGCGAGAAC | ACACCTGGAC | 19680 |
| AATTTAAGTT | GATAGAAACT | CCCCTGAACA  | GCTTCCTCTT  | GGTTTCCAAC | GTGATGCCCG | 19740 |
| AGGTCCAGCC | AATCTGCAGT | GGCCGGCCGG  | CCTTGGCGCC  | AGACTTTAGT | AATCTCCACT | 19800 |
| TGCTTAGACT | GGAGAAGCTC | CAGAGAGTCC  | TCGGGCAGGG  | TTTCGGGGCG | GCGGGTGAGG | 19860 |
| AAATCGCACT | GGACCCGTCT | CACSTAGAAA  | CACACGAAAA  | GGGCCAGGTG | TTCTACAACC | 19920 |
| ACTATGCTAC | CGAGGAGTGG | ACGTGGGCTT  | TGACTCTGAA  | TAAGGATGCG | CTCCTTCGGG | 19980 |
| AGGCTGTAGA | TGGCCTGTGT | GACCCCGGAA  | CTTGGAAGGG  | TCTTCTTCTT | GACGACCCCC | 20040 |
| TTCCGTTGCT | ATGGCTGCTG | TTCAACGGAC  | CCGCCTCTTT  | TTGTCCGGCC | GACTGTTGCC | 20100 |
| TGTACAAGCA | GCACTGCGGT | TACCCGGGCC  | CGGTGCTACT  | TCCAGGTCAC | ATGTACGCTC | 20160 |
| CCAAACGGGA | TCTTTTGTCT | TTCTGTAATC  | ATGCCCTGAA  | GTACACCAAG | TTTCTATACG | 20220 |
| GAGATTTTTT | CGGGACATGG | GCGGCGGCTT  | GCCGCCCGCC  | ATTCGCTACT | TCTCGGATAC | 20280 |
| AAAGGGTAGT | GAGTCAGATG | AAAATCATAG  | ATGCTTCCGA  | CACCTACATT | TCCCACACCT | 20340 |

|             |            |             |             |            |            |       |
|-------------|------------|-------------|-------------|------------|------------|-------|
| GCCTCTTGTG  | TCACATATAT | CAGCAAAATA  | GCATAATTGC  | GGGTGAGGGG | ACCCACGTGG | 20400 |
| GTGGAAATCCT | ACTGTTGAGT | GGAAAGGGGA  | CCCAGTATAT  | AACAGGCCAT | GTTCAGACCC | 20400 |
| AAAGGTGTCC  | AACACGGGC  | GACTATCTAA  | TCATCCCATC  | GTATGACATA | CCGGCGATCA | 20500 |
| TCACCATGAT  | CAAGGAGAA  | GGACTCAACC  | AACCTTAAAA  | GAGAGTTTAT | TAACTCGGCT | 20500 |
| CTGGAGGCCA  | ACATCAACAG | GAGGGCAGCT  | GTATCGCTAT  | TTGATCGTTT | TGGGGGTAGC | 20600 |
| AGCGCGGTGT  | TTGAGAAGCA | GTTTCAGGAC  | GCACAGCATG  | CCGTGAGGGC | CCACGGTSCA | 20700 |
| CTGAAGCGCG  | AAGCCGAGCT | CGGGACTCTG  | GTACGCAAGG  | CGGGCCAGAG | GTTTGAGGGC | 20700 |
| CTGAAAAGGG  | AACGGTCAAT | TTTGCGCCAG  | CCGCGCGACC  | TCCCACGGGT | CGCCGACATT | 20800 |
| GACGCGCTGT  | TCGACGCGCT | CGCGGACCTC  | AAAGAAGAGG  | TGGCGGTGCG | CCTAGATGCG | 20800 |
| CTGGAAGAGA  | ATGGAGAGGA | GACCCCCACT  | CACCTCTCTT  | CGGAGATCAA | GGACACAATC | 20900 |
| GTCAGGTGGA  | GGCTTGACGA | TTTGCCCCCG  | GTGTGCCCTG  | AACTCCCTA  | AGGCTACCCG | 21000 |
| GATTTTCAGAG | AGACCTTGCG | CCTCCACATG  | GCAGCTGAAT  | CAGCATATAC | AGGTGTCCAA | 21000 |
| GACTAAAAAG  | GCCACCGCGT | ATCTTAAAGC  | GCCCCGTGAA  | TGGGGGCGAG | GCACGACCCA | 21100 |
| GGATCCAGAC  | TGGTCCAAGC | GTCTGGGTGC  | TGGCGCTTTT  | GGCATAATCG | TCCCTATCTC | 21100 |
| CGAGGATCTG  | TGTGTGAAGC | AGTTTGATAG  | CGGCCGGGAG  | TTTTTCTACG | AGGCAATTGC | 21200 |
| CAACGACCTG  | ATGCAGGCCA | CCCGAGAGAG  | GTACCCCATG  | CATTCTGGTG | GATCTAGACT | 21300 |
| GCTAGGATTC  | GTGCAGCCTT | GCATACCCTG  | TAGATCGATT  | GTGTATCCTA | GAATGAAGTG | 21300 |
| CAACCTGCTG  | CAGCTGGACT | GGAGTCAGGT  | CAACCTGAGT  | GTCTATGGCG | CGGAGTTTAC | 21400 |
| CGGCTTAATG  | GCGCGGTGT  | CTTTTCTAAA  | CAGATACTGT  | GGCATGGTGC | ACTSCGACGT | 21400 |
| TAGTCCAGAC  | AATATTTTGG | CCACAGGAGA  | CCTAACGCCC  | ATGAACCCCG | GGAGGCTGGT | 21500 |
| CCTTACCGAT  | TTGGGTTCCG | TTGCGCTACA  | CTCTGGGAGC  | AAGTGGACTA | ACCTTGTGGT | 21600 |
| GACCTCTAAC  | CTGGGGTTTA | AGCAACACTG  | CTACGACTTC  | AGGGTGCCAC | CCAAACTCAT | 21600 |
| TTGTAAGCAT  | CTCTATAAGC | CGTCTTGCGT  | CTCTTTCCAG  | TGTTACCTAT | CCAGTCTCGG | 21700 |
| TAAGATGCAC  | GCGCAGGTAT | TGGACCAACC  | GTACCCCTATC | AGCCCTAACA | TGGGACTGAC | 21700 |
| CATCGACATG  | TCCTCGTTGG | GCTACACTCT  | GCTGACATGC  | CTGGAACCTT | ATCTCGATCT | 21800 |
| GCCGCTAAAC  | AACCTCTCTA | AGTTCTTGGG  | TTGAGCCACC  | AGAGACGGAC | GCCCCGAACC | 21900 |
| CATGTACTAC  | TTGGGCTTCA | TGATTCCCAG  | GGTGGTGATG  | ACTCAGATCC | TGTCCGCTGT | 21900 |
| GTGGACCATG  | ACGCTTGACC | TGGGACTAGA  | TTGCACCGGC  | AAAGCCCAGG | CGATTCCCAT | 22000 |
| GCGACAGGAG  | CACCAGCTGG | CGTTTCAGAA  | GCAGTGCTAT  | TTATATAAAG | CCAAACAAAA | 22000 |
| GGCAGAGTCG  | TTAGCGAACT | GCTCCGATAA  | GCTAAACTGC  | CCCATGTTAA | AGTCTCTCGT | 22100 |
| TAGAAAGCTA  | CTAGAGCGAG | ACTTTTTTCAA | CCATGGAGGC  | CACCCCCACA | CCCGCGGACT | 22200 |
| TGTTTTCTGA  | AGACTATCTG | GTTGACACCC  | TGGATGGGTT  | AACAGTGGAT | GACCAACAGG | 22200 |
| CTGTCTCTGC  | AAGCTTGAGC | TTTTCAAGT   | TTCTAAAGCA  | CGCCAAGGTT | CGAGACTGGT | 22300 |
| GCGCACAGGC  | CAAGATCCAA | CCCAGCATGC  | CTGCGCTGCG  | CATGGCTTAC | AACTATTTCC | 22300 |



|            |            |            |             |            |            |       |
|------------|------------|------------|-------------|------------|------------|-------|
| TTTTTTCAAA | AGTGGGCGAG | TTTATTGGTA | GTGAGGATGT  | GTGTAACCTT | TTGGTGGACC | 22440 |
| GTGTGTTTGG | TGGTGTCAAG | TTACTGGAGG | TGGCCAGCGT  | GTACGCCGCG | TGTTGGCAAA | 22500 |
| TGAACGGACA | TCAGCGGCAC | CACATCTGCT | GTCTAGTGGG  | GAGGGCCACT | AGTAGTCAGA | 22560 |
| GTCTGAACCC | CGTGTGGGAC | GCCCTGGGAG | ACGGAATTAT  | ATCTTCATCC | AAGTTTCACT | 22620 |
| GGGCAGTTAA | ACAACAGAA  | ACTTCAAAAA | AGATATTTCAG | CCCATGGCCT | ATAACGAACA | 22680 |
| ACCACTTTGT | CGCGGGCCCC | CTTGCCCTTG | GGCTGCGGTG  | CGAGGAGGTG | GTGAAAACGT | 22740 |
| TGCTGGCCAC | CCTTTTGCAC | CGGACGAGA  | CAAATTGTCT  | CGATTATGGG | TTTATGCAGA | 22800 |
| GTCCGCAAAA | TGGAATATTT | GCGGTGTGCG | TGGATTTCGC  | GGCGAACGTC | AAAACGACA  | 22860 |
| CCGAGGGTGG | TCTACAGTTT | GACCCTAAC  | GTAAAGTGTA  | TGAAATAAAA | TGCAGGTTCA | 22920 |
| AGTACACCTT | TGCGAAAATG | GAGTGTGACC | CCATATACGC  | CGCGTATCAG | CGGCTGTACG | 22980 |
| AGGCACCCGG | AAAGCTGGCA | CTGAAGGACT | TCTTCTATAG  | CATTTCCAAG | CCTGCGGTGG | 23040 |
| AGTACGTGGG | ACTTGGAAAA | CTGCCCAGTG | AATCTGATTA  | CTTGGTGGCT | TATGATCAGG | 23100 |
| AATGGGAGGC | GTGTCTCTGC | AAAAAGAGGA | AATTAACGCC  | CCTTCACAAT | CTTATTAGGG | 23160 |
| AGTGTATTTT | GCACAACTCG | ACCACGGAGT | CTGACGTCTA  | CGTACTTACT | GATCCTCAAG | 23220 |
| ATACTCGGGG | TCAAATCAGT | ATTAAAGCCC | GCTTCAAAGC  | CAACCTCTTC | GTGAACGTCC | 23280 |
| GTCACAGCTA | CTTTTATCAG | GTATTGCTGC | AGAGTTCGAT  | CGTCGAGGAG | TACATTGGCC | 23340 |
| TAGATAGCGG | CATTCTCTGC | CTCGGATCAC | CGAAATACTA  | CATCGCCACC | GGCTTCTTCA | 23400 |
| GAAAGCGGGG | CTATCAGGAT | CCTGTCAACT | GTACCATCGG  | TGGCGATGCT | TTAGACCCCG | 23460 |
| ACGTGGAGAT | TCTTACGCTG | CTAATCGTAA | CCCCCGTCTA  | CTTTCCCGCA | GGCGCAAAGC | 23520 |
| ATCSTCTGCT | TCACCAAGCT | GCCAACTTTT | GGTCAAGAAG  | TGCGAAGGAC | ACCTTTCCAT | 23580 |
| ATATCAAATG | GGATTTCTCC | TATCTATCTG | CAACGTCGCC  | TCACAGCCCG | TAGACGTGGA | 23640 |
| CGGGGAACCG | CTCGACGTAG | TGCTGGACTA | TGACCCCAT   | CGCGTTTCAG | AAAAGGGCAT | 23700 |
| GTGCTTGAG  | CAATCGCAAT | CCCCATATCC | CGCATTAAAA  | AAGAAGAAAA | AAAATAAAGA | 23760 |
| AGCAATTTAT | TAAGCAAACA | GTATGGTTTT | CTGTACGTAT  | TTTATTCGGT | GGTGGGTGAA | 23820 |
| AAATAACGGG | GGATGGAGGA | AGAGGGATGG | GTTTATAATG  | CCAATATATC | AGCTAAATGA | 23880 |
| ATATCATTTG | CGTTTCTGTC | ATTTCACTGT | CACTTTCATG  | GTCCGACTGG | TATTGGGTCC | 23940 |
| TCGGGGCGGG | CGTCGATATG | TCCTTCACTT | TGGCGCGGGC  | TCTGGTCTTT | GCTGGGAGGG | 24000 |
| GCGGCGGTTT | CTGGTGAACA | GTCCGAGTTC | TATCGACCGT  | CGGCGCCGAC | GTCCGCCAGG | 24060 |
| GCATGTATGC | CGCACTCGGC | GTACAGAGTC | CCCAGTCGCT  | CCTTATAACG | CGTATAACGA | 24120 |
| TGGCTAGGAT | GCACAGTATA | GGGATACAGG | AGATATTGAT  | AGCCACTATG | TAGTGGAGAT | 24180 |
| TAGCCTGCAC | GAACGCGTTT | TCATACCTGA | TGACAGGCAG  | CAGTAGAATC | AGATAACCCA | 24240 |
| CCAATACTCC | CACGTAAAAG | CCTACCTGCC | GTCTCATAAA  | CTTTACCAGG | AAAAATTCCG | 24300 |
| TGTTTATGTA | CCACACGACC | GTCAAGGCTA | GGAACATGTT  | CACCGCACCA | AAAATGGCGT | 24360 |
| CTGACACGAG | CACGTAAAAG | CTGTTGCCAA | CGGCCATCAT  | GGTGCTCAAT | GAAAACAGCA | 24420 |

|             |             |            |            |            |            |       |
|-------------|-------------|------------|------------|------------|------------|-------|
| GCATTTCCAA  | GGCGGTTGTT  | GATAGGTACA | GGTTGACGCA | GACCGGTTTC | CACCGAGTCA | 24480 |
| GCAGTGACTC  | CATCATGGTA  | TTATCAGGTA | CGTGCTGTTT | CAGGAGAGGT | ATTTCCCACT | 24540 |
| GGCGGAGGTT  | ACATGTTATC  | AGTGACTGGA | TGTGGGCAAA | GGATATGCAA | AAATGAATGC | 24600 |
| AGTAGACAAA  | GGCTGCCATA  | AGTACGTGTT | TATATGACAG | AACATGGATA | AACAGTTGCA | 24660 |
| TGCTCCACAT  | CCTTAAGATG  | GCGACATAAA | GCACGCTATG | TGATCCAAAT | AGCGCTATCC | 24720 |
| AGGATTGCAT  | GCTCATCATG  | GTAGTGGCGT | GAACATGCTT | GGCCCGATAT | ACGCGCACCG | 24780 |
| CCGCGAGACA  | GTAGTATACT  | ATGGCAATGC | CGTCCACGAT | AAAAGTCCAA | AATATGTACA | 24840 |
| CCAGCATCTC  | TGGTTTCTCT  | AAAAACAGGG | TCGGGGTGAG | GTGCTTCGGT | GAGTTGCGCA | 24900 |
| CCGTGAGGTT  | TAGCGCGCTG  | TAGTTTACCA | GATTGTTGAA | GTAGCAGGGG | AAACCAAGGC | 24960 |
| CCTCGTACGT  | GGCGGCCATG  | GGCAGCACTG | CAGAGCAAAT | GTACATAATT | ACAGCCACAA | 25020 |
| ACACACAGCTT | GACCCAGGAG  | GACATGAGAA | AACGGTCGCT | CTTTGAAGCG | CGCATGTTTT | 25080 |
| TCGGTCTTTT  | TAACTTTTCG  | CAGGCGGCGC | TGCGGCGGGA | GAGCCAACTT | GATGCCACTG | 25140 |
| CCTATCGCGG  | TTGACTTTTA  | AATACGCGCC | CCGGGCAGAA | GCCAGAGGTA | GTGCACTCAT | 25200 |
| TGACTCAATG  | GCAACGAGCG  | AAGAAACGGC | GGCCGGTTAT | GTGATCGGTG | TCTACTTTCA | 25260 |
| CAGCGTTTAC  | GTCCACTGCC  | GCATTATTGT | CTGGCAGGTT | AATTTTCTAC | CCCTGGACCC | 25320 |
| AAACGACGGG  | GAGACTGAAT  | GCTACTTTGT | GGTGGACACG | CTGACGAAAG | AGGCGATGGA | 25380 |
| GGCGATGCCC  | GAAATCCAGG  | AATGCGTCCC | GTCTATTACT | GAACACGCCC | GTGACTGTGC | 25440 |
| GATCTGGGAG  | TTGGCGCTGC  | GACTGCAGAA | TCAGACGATC | GTCAAGGCCG | TCGCGACAGT | 25500 |
| GTGCTTTCCG  | GTGTTTCTAA  | TTATGACTGT | GGGTGCGATA | GTGAATGATG | TGATTCCCTG | 25560 |
| CCCCAACGTC  | AGAACACCCA  | GACCACTAGC | CTGTGCTTAC | CTACACTGTG | AGGCGACGGT | 25620 |
| GACCTTTGAG  | GTCCCACTAA  | CCGGGCCCCG | GGCGTCCACC | GGAACGTGCG | ACAGCTCTAT | 25680 |
| CTATAGGGAA  | TGTGCGATCT  | CGGCTATCGA | GATATGCTTG | AAGACCACTC | GAGGCATATA | 25740 |
| CTCCTGCCAG  | TCGAACGAGG  | CCCTTGAGGC | CAAGAGGGAA | AACCGAGGTT | TAGACATATC | 25800 |
| AGATGTGTTT  | GTCTGTCTCA  | CGTATGATAT | CCCTATCGCA | GGGCGGGTCC | TTTCTCTGCT | 25860 |
| GGTGCCCCAC  | GCGCCCCGCTT | TTGACGTCTT | ATGGATCAAT | GAGGACAGCA | AGTGGAACTG | 25920 |
| GGCAGCCGTC  | GAATTTTTCA  | GAGCCCTACA | CCATAAGCTG | TTCAGTGAAC | GCAATGGTAT | 25980 |
| ACCCCTCTCT  | TGGTTGTACG  | TGTTCCCGGG | AGCTGTGGAA | GAGGGCACAG | CCTTTTCGCT | 26040 |
| ATTACTTTCC  | GCATTCCCTT  | GCATACCTTT | GCGGTATGGG | TGCGCTACCT | CTCTGGACAG | 26100 |
| GGCTTCGCTG  | CAGTGGGACC  | TATTTGAACC | GCACATCCTG | ACCCACTTTG | ACGGGATAAA | 26160 |
| GCGAACTTCT  | TTGGCAGATA  | CAGTGTGTTG | CTACGACTCC | CTGGCCATTT | CAAGGGAATG | 26220 |
| TGAAGATCAG  | TATGTGTGGC  | CCACGCTCTG | CACGTGACAT | AATATTAAAT | TGTGCACGGA | 26280 |
| TAGTGACACT  | ATGGCCATCG  | TTAGAGAAAC | ATCCGGTCTG | GTGGCCCTGA | ATCTAGAAGC | 26340 |
| CCTGTTGCGC  | ACCGACTCCG  | TATTATCGCG | GGTCTCGTCC | ATTGTCTCAC | TCGATACGCT | 26400 |
| CTTGGACCTT  | TCCACCCCGG  | AGTSCCGTAG | GAGCGTGGAG | CTTAGATACA | ACTCACTTTT | 26460 |

|             |             |            |            |            |             |       |
|-------------|-------------|------------|------------|------------|-------------|-------|
| GTCGACTGTA  | TTATCATGGT  | CCACCTCTAG | GGGTCCACAA | TGGGCGGCAA | TGGTGAAGTG  | 26520 |
| GAAGTTATTT  | TTCTCTGTCC  | AAGCTTTGGA | GCCTGAGGTG | AGACCTACTG | TCCCTGCTTG  | 26580 |
| AAGCGGAGAG  | GGGGTGGTGC  | GAGTTGGCAG | TTGACGGGTT | TGTGATAGCT | GGAGTGTCTG  | 26640 |
| CCACGGCACA  | GGACCCATTA  | ACTTTCTTAT | GTGTTTATTT | TTAGCAATGG | TCTCCAGAAT  | 26700 |
| TCAAGGATCT  | CAAAAGGGCC  | TGCCAGATGG | CCGGGTTTAC | TCTGAAGGGG | GGGACTTCCG  | 26760 |
| GGGATCTTGT  | ATTCTCATCG  | CATGCGAACT | TGCTCTTTTC | AACCTCGATG | GGATATTTCC  | 26820 |
| TCCATGCAGG  | CAGTCCAAGG  | TCGACAGCGG | GGACGGGGGG | TGAGCCTAAC | CCACGTCCACA | 26880 |
| TCACCGGACC  | AGACACTGAG  | GGAAATGGGG | AACACAGAAA | CTCCCCCAAC | CTCTGCGGGT  | 26940 |
| TTGTTACCTG  | GCTGCAAAGC  | TTAACCACAT | GCATTGAACG | AGCCCTAAAC | ATGCCTCCCG  | 27000 |
| ACACTTCCTG  | GCTGCAGCTG  | ATAGAGGAAG | TGATACCCCT | GTATTTTCAT | AGGCGAAGAC  | 27060 |
| AAACATCATT  | CTGGCTCATC  | CCCCATTCGC | ACTGTGAAGG | GATCCCGAGT | TGCCCCCTTT  | 27120 |
| TACCATTTGA  | CTGCCTAGCA  | CCAAGGCTGT | TTATAGTAAC | AAAGTCCGGA | CCCATGTGTT  | 27180 |
| ACCGGGCAGG  | CTTTTCGCTT  | CCTGTGGATG | TTAATTACCT | GTTCTATTTA | GAGCAGACTC  | 27240 |
| TGAAAGCTGT  | CCGGCAAGTT  | AGCCACAGG  | AACACAACCC | CCAAGACGCA | AAGGAAATGA  | 27300 |
| CTCTACAGCT  | AGAGGCTTGG  | ACCAGGCTTT | TATCTTTATT | TTGAAAAAAG | GGAAACAATG  | 27360 |
| GGGGGTTTGA  | AAAGGGTGCA  | CATTTTCAGA | TATTTTAAAA | CTTCATTGTT | CTCCAGGTGC  | 27420 |
| TTGGTAAAGA  | TGSTATCACA  | ATAAAAAATG | TTTACTGGGT | CCGCGCAGGT | TTGTTTGTCA  | 27480 |
| TCTTCATTCT  | CTCCACTAGA  | CTCCAGTTTA | AAAGACTCTA | GATAAATGGG | TTTCATTAGT  | 27540 |
| CCCCCATGCG  | GGGTTGAAGC  | GTCGCCTATC | GCCTTATGAA | GCTTAAACAT | AACGAGTGGG  | 27600 |
| GTGGCCCTGA  | AATGATCGTC  | CACGGACAGC | TCGTAAACAA | AGGCGGCCGT | GGCAGTCAAC  | 27660 |
| GTCTCTATAC  | GCTGCATGAC  | GAAGGCCGGG | TCCATCCCCG | GCGTCTCTTC | ATGTGTCTTT  | 27720 |
| CTGGCGCGAC  | AAATAATAGA  | TCTCAAAAAC | GTTGGTGACA | TGTCTCGACA | GTTCTCGAGC  | 27780 |
| ATCGATAACA  | GGCAGCAGAG  | CTCGGTTATG | CCGCGAGATG | TAGGTCTAAG | GAGGCACACT  | 27840 |
| CGCTCTTGGA  | ACACGTGAGG  | GTGTAGGTCT | ATGTGGGTCA | CCATGTCTTC | GTGCTCCACC  | 27900 |
| AGGCACACCA  | CCGTAAATCC  | CACAAAGTTG | GGCGAGGACA | GGCGAGATTT | CACGTGCTCC  | 27960 |
| CTGAGACACG  | CTATATCTAA  | GTGGCCCATC | ACGGACATTT | TGGGGGTATT | GCTTCCCAAC  | 28020 |
| AGTGCGTTGT  | TTTTCTTATG  | CACTTCCAGG | ACAAGGCCGG | GCACCACAGG | GTGGGGGTAT  | 28080 |
| ACGGGACAGG  | CCTCTTTCTGA | CTCGCGAGTC | TTGGGGGCAT | GAGTACTCAT | TGGCACTCCA  | 28140 |
| GTCACTCTCG  | CCAGGGCCCT  | TTCCAGGGAC | ATTCTCGAAG | GSTGGTGTAA | CTAGACAGTA  | 28200 |
| TTTTCTGTCC  | ACCTCGGTTA  | TATACACAAA | GAGTCTGCTA | GTCTGATATA | AATAGGCCCG  | 28260 |
| GATGTCTCTG  | AAGCTGGAGG  | ATACGAAGGA | GTGACTAATG | AGCTCCATCT | GAAGCAGGTC  | 28320 |
| CGCGATCACA  | TACGTGAATG  | GACCAAGCAG | GATGGATATG | GTGTCTTSAG | AATAGGTGAC  | 28380 |
| GCTGAGCCCG  | TGCCCTTGGT  | TCTCAACAAC | GGGAGCCAGC | TTGTAGGTTT | GAAACATCTC  | 28440 |
| GCTTTCCCAAC | AGGTTCTGTG  | GATCTTTCAT | GCTTTCTCTC | ACTGGGGGTA | TGTAAGAAGA  | 28500 |

|            |            |             |            |            |            |       |
|------------|------------|-------------|------------|------------|------------|-------|
| GAAAAAGCTA | TTTAGCACGG | CACTGCCCCGA | TGGGATATGG | GAAGACGTTA | GCTGCAGAGA | 28580 |
| GGGGTCCTGT | AAACGTCCCA | GAGATTGAAA  | TGTGTTGGCG | GTCAGCAGAT | TCACACTCCC | 28600 |
| GGGACCCCTT | GCGTCACCGG | GCTGTTGGTG  | TGACAGCTGT | GTCTCAATAC | ATTTTAGCCT | 28620 |
| CTTCATGCAG | AGCTCCCTCT | CCTTTTCAAG  | TTGAGTTATT | GTGTCAAATT | GTTCGTTTAT | 28640 |
| CTGSTTGGTG | AGACACTTGA | AAACGCTGTT  | GGACACCTGG | CGCCTGASCC | CCTGASTGGT | 28660 |
| CGTCTCTTGG | CCTGTGCCGA | ATAGTTTATT  | CTTGTCTACT | ATGTTTTTGG | ACACGTGGGT | 28680 |
| GACAAAGTCC | TCCACGACGT | CGGTGACACC  | GCTCACTGTC | TTGTTTTCTG | CCAGTTTCAT | 28700 |
| GAGCAGGTG  | AGGAGCTCTC | GCTTGGGGTC  | TGTTCTCTGA | GAGGCCTGCT | CCAGGTGGGT | 28720 |
| CATGATGTCT | TTGTACACAT | TGTTACAGGC  | GCTTCCAACG | AGGGCCTTGG | TGGGGGCTGT | 28740 |
| GTTCAGGAGC | TGGCAAAGTT | TTGCGTGCTC  | TGCCGTCCGG | TGACAGCTCA | TAATGCTGGT | 28760 |
| ATACATCCTC | TGAATGGGGC | TGTCAAAGAT  | CACCCGCCCA | GCCAAGATGG | CGGGCATAGT | 28780 |
| AATCACCTCC | ACATGAACCC | TTTTCTGCCT  | ATACAATCCC | ACGAAAGTGT | TTTTAACACA | 28800 |
| GTCATAGTCT | ATGCTCACCT | CTGAGTAGCC  | CGGAATATAG | AGGGCGCTTA | AAGTAGACAC | 28820 |
| CAGGTTGCTA | ATCTCCTGAG | TCACGCTGGT  | GAGTATCCGG | CCTATGGTTT | TTTCACCAGA | 28840 |
| GGCCAGACGC | TGGCAATCTT | TCATCAGCTG  | TTCTGGGATA | GAGTTAACCA | GCTTGTGGTC | 28860 |
| GGGTGTGTGC | TTGACGACTG | GTACCATTCG  | TACCGTGACC | ACCCAGTCTA | CCTATCTCTC | 28880 |
| ATACGAGAGC | TGTGTCTTGG | CGTAGAGGAC  | CCGCTTGATG | GCATTGAGAA | GCAGGTGGTC | 28900 |
| TAATGTGATG | CGCATAGTCT | GGSCCCAGGA  | GTGGAAGGTT | GACCTTCTGT | AAGACCCCCA | 28920 |
| CTGTGCTTCC | TTTTCTGGGC | ACCTGGTTTT  | TGCTGAGGAC | TCGTATGTCC | TCCAGTCGGA | 28940 |
| CAAGACGTGG | TCGTAGCTAC | AGTTGGCCAA  | TGCATTCTTG | TACAGGTGGA | TAAATAGCTG | 28960 |
| TCTGAAAAAA | ACACCCGGGT | TTGCGAGGCT  | GCAGTGTAGA | GTCTGACCTC | TGACATAAGA | 28980 |
| ATACTTGCTT | TGCAGGATCT | CAAAGAGGGA  | GATGGACAGC | TGGGAAGGGT | GCACTGATAT | 29000 |
| GGACGAGCCC | AGCCCGGGGT | TCATCCTCAA  | CATGACATCG | GATGCCAAAG | TCAGGAGCGT | 29020 |
| AGTGGAAACG | ATTGACAGGT | TGTCAAATAT  | CACTACCTCG | CCCCCGSAGA | TGGGCTGGTA | 29040 |
| TGACCTAGAG | TTGATCCAC  | TGGAAGACGA  | AGGCCCTTTT | CTGCCSTTTT | CGGATACGCT | 29060 |
| AATAACGGGG | ACTGCAGGAG | CGGGGAAAAA  | CACCAGCGTA | TCCGCCCTAC | ATCAGAACTC | 29080 |
| CAACTGCCTA | ATTACGGGGG | CTACAGTGGT  | AGCGGCACAG | AATCTTTTCA | GGGCTTTAAA | 29100 |
| GTCTTACTGT | CCCACTATAT | ACCACGCCTT  | CGGATTCAAG | AGCAGACACA | TTAATATCTG | 29120 |
| CCAGAGGAAA | GTGCCCAAGG | TAACTCAGTC  | CTCCATCGAG | CAACTCCAGA | GATACGAGCT | 29140 |
| GGCTAGGTAC | TGGCCAACTG | TCACCGATAT  | TATTGAGAAA | TTTATGCGCA | AGAAACAAA  | 29160 |
| GGGGCACTAT | AGCTCCCTCT | CTCAAAGCGC  | TTTCAGACTC | CTTTGGCCTA | TGGGTGGAGC | 29180 |
| CAATTTGTGG | ACGAGTAACA | TTATCGTGAT  | AGACGAAGCT | GGAACCTCTT | CCTCCCATAT | 29200 |
| TTTGACGGCC | GTGGTGTCTT | TCTATTGGTT  | TTACAACAGT | TGCTTGGACA | CCCCGCTATA | 29220 |
| CAGAAATGGT | GCCGTGCCTT | GCATAGTCTG  | CGTGGGGTCT | CCCACCCAGA | CGGACGCTTT | 29240 |

|            |             |             |            |             |            |       |
|------------|-------------|-------------|------------|-------------|------------|-------|
| TCAGTCGGTC | TTCAACCACA  | CGCAGCAGAG  | AAACGAGATA | TCTGCCTGTG  | ATAATGTGCT | 30600 |
| CACCTTCCTA | TTGGGAAAAC  | GTGAGGTTGC  | AGATTATATT | AGGCTGGACG  | AGAATTGGGC | 30650 |
| CCTATTTATA | AACAATAAGC  | GCTGTACGGA  | TCCCCAGTTT | GGTCACTTGC  | TGAAGACCTT | 30700 |
| AGAATATAAT | CTAGACATAT  | CACCAGAGTT  | AATGGACTAT | ATAGATAGGT  | TTGTGCTTCC | 30750 |
| GAAGAGTAAG | ATTCTGGACC  | CGCTCGAGTA  | TGCAGGGTGG | ACAAGACTCT  | TCATCTCACA | 30800 |
| CCAGGAGGTG | AAGTCTTTTC  | TGGCAACGCT  | GCACACCTGC | CTGTCTGAGTA | ATAAGGATGC | 30900 |
| TGTGTCCACA | AAGCTTTTCA  | CCTGCCCCAGT | GGTCTGTGAG | GTGTTTACAG  | AGCCATTTGA | 30960 |
| GGAGTACAAA | CGGGCGGTAG  | GCCTCACACA  | CATGACTCCC | ATAGAAATGGG | TAACAAAAAA | 31020 |
| TCTTTTCAGG | CTAAGTAAGT  | ACTCGCAGTT  | TGCTGATCAG | GACATGGCTG  | TGGTTGGGAC | 31080 |
| CTATATCACA | GACGCGTCCA  | CACAGATCAC  | CTTCGCCACT | AAATTTGTCA  | AAAACAGCTA | 31140 |
| TGCTACCCTT | ACTGGAAGA   | CCAAAAATG   | TATATGCGGG | TTTCACGGGT  | CATACCAAAG | 31200 |
| ATTCAAGTCC | ATCCTAGACG  | GGGAGCTATT  | TATCGAAAGT | CATTCTGCACG | ATAACCCCGC | 31260 |
| TTATGTGTAC | AGTTTCCTTA  | GTACCTGTCT  | ATATAATGCC | ATGTACTCAT  | TTTACGCGCA | 31320 |
| CGGGGTGAAG | CAGGGGCATG  | AAGAATTCCT  | CAGGGACCTC | AGGGAAGTGC  | CGGTGTCTCA | 31380 |
| AGAGCTGATC | TCTGAGATGA  | GCTCCGAGGA  | CGTTCTGGGG | CAGGAGGGGG  | ACACAGATGC | 31440 |
| CTTCTACCTC | ACCGCCAGCC  | TCCCACCATC  | CCCCACCCAC | GCGGCTCTTC  | CAACACTGGT | 31500 |
| GGCCTATTAC | TCCGGGGCCA  | AGGAACTATT  | CTGCAACAGG | CTGGCCCTGG  | CACGCCGACA | 31560 |
| CTTTGGTGAC | GAGTTCTCTC  | ACTCCGATTT  | TTCAACGTTT | ACGGTGAACA  | TCGTGGTGCG | 31620 |
| AGATGGCCTG | GACTTTGTGT  | CCACTTCCCC  | CGGGCTCCAC | GGTCTAGTGG  | CATACGCATC | 31680 |
| CACATAGAC  | ACCTATATAA  | TCCAGGGATA  | TACGTTCCCT | CCAGTGAGAT  | TCCGCCGTCC | 31740 |
| AGGAGGACAG | CGCCTCAGCG  | AGGACCTGCG  | CAGAAAGATG | CCCTCCATAG  | TTGTCCAGGA | 31800 |
| CTCATCGGGG | TTCAATTGCCT | GCCTGGAAAA  | TAACGTCAAC | AAGATGACAG  | AGACCTCGA  | 31860 |
| AGGTGGCGAC | GTGTTTAACA  | TATGTTGTGC  | AGGGGACTAC | GGTATCAATT  | CTAATCTGGC | 31920 |
| TATGACCATA | GTGAAGGCAC  | AGGGGGTTTC  | ACTAAGTAGG | GTGGCCATAT  | CGTTCCGCCA | 31980 |
| CCACCGCAAT | ATCAGAGCCA  | GTCTAGTGTA  | TGTGGGTGTA | TCCAGGGCCA  | TGCACGCTCG | 32040 |
| TTACCTGGTA | ATGGACAGTA  | ATCCCTTTAA  | GCTAATGGAC | CGCGGTGACG  | CCCAGTCCCC | 32100 |
| ATCCTCAAAG | TACATCATCA  | AAGCCCTATG  | CAACCCCAAG | ACTACTCTGA  | TCTACTGACC | 32160 |
| CGTACCCCTC | TCTTAGGACA  | CTGATGTGTT  | TGGGAATATA | GCATGAGACT  | TGACACCTAT | 32220 |
| AATGGTCTGT | ATTGACACCA  | TTCTTTTATT  | TATCAGTCCA | GCCACGGCCA  | GTTATATGCA | 32280 |
| CCGTTTCCAC | ACAGGGGTGG  | CGTGGAGGCC  | AGGATGCGGG | TTGGGTCTGT  | GCACCTGGAC | 32340 |
| CCCGCGGTAG | TTGTGCTTCC  | TGATGAATTC  | GAGTGGGCGG | AAGTACTGGG  | AGATTGGGTT | 32400 |
| GGGAGGTGAC | CCTTTGTGCT  | CGACGGGAGC  | ACGATCACCG | TCACGGCGGA  | CGAGGGCTCC | 32460 |
| TGGTCTGTGT | CAGTCCCGCA  | GGATATAATT  | ATCACGGAGC | CCACTGCTTT  | GCGGCTTAAG | 32520 |
| TTTGGTTGTC | TCTGGCAGCG  | CACCACTCC   | TGGCTACCAG | AGGAGGCGGT  | AGACTGCCTT | 32580 |

|            |            |             |             |             |             |       |
|------------|------------|-------------|-------------|-------------|-------------|-------|
| TTGCGCTTCT | GGCCACGTC  | CATGAGCCCG  | ATTCTCTGAC  | TCAATACTTC  | CCCTTGGTCT  | 32640 |
| TCTCCGTCT  | CCTCGGACGA | GGGTGGCTGG  | TGGGAAAAAT  | GGCGGGGGTC  | GGTAAACGGG  | 32700 |
| GCCTCATTGT | TCACSTCCGG | AGAGTTGGAA  | CTGTCACTGC  | TATCAGAGTC  | CGATGTGAGG  | 32760 |
| TCGACGATCG | CGGTGGGTGC | GGCGCGCAGG  | GGGCGCCACG  | AGGGCCCTTC  | ATCAGGGTCC  | 32820 |
| CTGTATGGTG | AACTTTGTGT | TCCAGGTACA  | CTATTTCTGG  | AAGCAGGTGA  | AAGTCCGTAT  | 32880 |
| GGCCCGGTCC | CAGTGTATGC | CGCCATCGGT  | TCCAGGATAG  | CAACCCCTTC  | GTCTCTGAA   | 32940 |
| GGTGAGAGCC | CAGCAGGGGA | AAATCCGTCA  | TCCTGACTAA  | CCCATCCCAT  | GGACGGCTCG  | 33000 |
| GACTCCGGCG | TGTCCGTGA  | ACTGCGCAGG  | CGGCCCGCTA  | CCACTGCTAC  | CGGTTTGGGC  | 33060 |
| GTATGGGCCC | GTCTGSCCAG | AGGCCTCGGG  | CGCAAGTGAG  | ATAAAGSTTG  | AAAAAAGTCT  | 33120 |
| GCAGGGTACC | CCTCTGGCTC | GTCTTCCTCC  | TGAACATCGT  | CATTTTCTTC  | TTCATCTTCA  | 33180 |
| TCTTCCTCAT | CCTCGTCATA | TTGAGATTCC  | CCGCTCGACT  | GATCCGGGGA  | TATCTGTAGA  | 33240 |
| TCCAGAGGGG | TTGCTGGCGG | CGATGGCGTG  | TCCTCGGCCA  | AGACGTCTGC  | TGGGGCAGAC  | 33300 |
| ATATCTATCA | CCGTGGGTCC | AGCATAGCCG  | CGCGCCCTGC  | CAATTCCTGG  | AAGTGTATGA  | 33360 |
| AGAGSTGGAG | GTGGGAATAT | GAACTTCACG  | GGGGGTCTGC  | TGGGAGGGCG  | TCCTTCAATT  | 33420 |
| GGAAAGCATT | TCTCTTCATC | GTGTGTGCTA  | GACGAGGTCC  | TCACAAACAT  | CGCCATGGCC  | 33480 |
| TTGTACGGGG | TTGACCGCTA | GGGGCGGAAA  | TTTACAAAGC  | ACACGAGTTA  | TTGCCTTTAC  | 33540 |
| TGCTCCAAAC | GGCCCCAGTC | CACAGTCTCA  | CGCCGGTGGC  | GAGTCAAATA  | GTCTTTGGCT  | 33600 |
| AGSTTAAAGT | GATTACAGCC | CTGGAACCGA  | GGCCATCGCG  | AGTGTCCGGC  | ACCAAGAGAG  | 33660 |
| GGCAGCGGAG | ATGGATGCTG | GGCCGTAAGC  | ACCAGGTGTI  | TCTGTGCTTT  | TATGAGCGGA  | 33720 |
| GTTCTGTCAA | TGGCCCTTGC | CCCCCACAGG  | AGAAAAACGC  | AATGTTCTAA  | CTTTGAGGAT  | 33780 |
| ATGCTACTGA | TGATGAAACT | CGTGAACCAA  | TCCAGGCCAA  | GTCCCTCTGT  | TGAGCCGGCC  | 33840 |
| CTCCCTTTCT | CCACCGTCAA | AACTGTGTTT  | AGTAGCAACA  | CACCCCTGGG  | AGCCACAGCTG | 33900 |
| TCGAGGCACC | CGTGGGAAGG | AGTACTGAAA  | TTGGGGAGCG  | AAGCCTCTAG  | CTCTCTAAAG  | 33960 |
| ATGCTTCTCA | AACTGGGTGG | AACTGTGACAT | TGCGGATCCA  | CACCTAACGC  | CAGGCCAGTA  | 34020 |
| GCTTGGCCCT | TGTGGTACGG | GTCTTGCCCT  | AAGATCACCA  | CTTTAATATC  | CTCTGGATCG  | 34080 |
| CAGCAGTGGG | ACCACCACAT | CAGCTTGTCC  | TGTGGGGGAT  | ACACTGTGCT  | GGTTAGCCTA  | 34140 |
| AGTTCCCGAA | TCTGTCTGAG | CAGCGAGAGC  | AGTTTCTGTT  | TCAGAAATGA  | TGAGAGGCTC  | 34200 |
| AGAAAGGAAA | TCCACTTAGG | TGCCAGTAAC  | AGATCCCGGT  | CGTCCACCCC  | CTGACTGATG  | 34260 |
| GATAGGGTGC | CCCTAAAGAC | CGTCTGTTGC  | AACCATGCGT  | CCATGTTGAA  | CTTATTTTCC  | 34320 |
| CTTTTGACCT | GGGTGCGCTC | TCCGGCTGCT  | GCTTTTAGCC  | CGAGTCTGAC  | TTCCCTAAC   | 34380 |
| AGAACCTGTC | CGGTTGATGG | CTTTTCCCAC  | GCTTATTATA  | ATTATGTTTA  | CGTTGTGAAT  | 34440 |
| AGAGCTATCT | GCAGTGGTCC | CGTTAAAAAC  | TACAGTATAG  | GGCGTCAAAAC | TTCTTTGTAA  | 34500 |
| ATACCACAAC | AACCTCAGGT | TTTCTGTGGA  | CGCCACAGGAC | CCCAATCTTC  | GAACGACCGC  | 34560 |
| GACTAAAAAT | GACCTCAGAT | TAAACCCATT  | CACGCATGTT  | TCCACGGTAA  | TCTCGCCTGT  | 34620 |

|            |            |            |            |             |            |       |
|------------|------------|------------|------------|-------------|------------|-------|
| TTTGCTTCGC | AGCTTGGCTA | TACAGACCCC | GTTGCAGTGA | TTCCGGATCGG | CGAAGTGGAT | 34680 |
| AGAGTGGACC | GCAAAGAACA | ACGGCAGGGT | AGAGGCTGCC | GATGCCTGAA  | TTGCGCAACA | 34740 |
| TGSTAAGGCG | ACGTATGCGT | GAGATGTGAC | CAATAGGGTG | GTCCACAGGA  | CGGCAATAG  | 34800 |
| CGCAAAGATC | CCCATGGGGC | AAATCCGGGT | TTACCCCTTG | TGTTGCCTGG  | TTCCGTGCTC | 34860 |
| CCCAGGGAGC | CCCCTTCCGT | AATATCTGTT | TTATATAGTG | AGGGTTCACG  | CATGCGCGAG | 34920 |
| TCCCCACTAA | TGAGGACAAT | TACTGAAATT | GACCTTTTCG | CGACACGGGG  | GTGAGGTCTA | 34980 |
| TTTCCCACGA | CATACTTCCG | CGGAAAAATA | CCCACGCTCC | TTAATTTCCG  | TGGGAAGACC | 35040 |
| ATGGGGGAAA | TGTGGCATT  | CCTGACACGG | TTTCAATCAT | ACTCATCGTC  | GGAGCTGTCA | 35100 |

## (2) INFORMATION FOR SEQ ID NO:19:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35100 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:19:

|            |            |            |            |            |            |      |
|------------|------------|------------|------------|------------|------------|------|
| CACGTCTGGC | TGAGATTTTC | TAAAAAGTCA | TCCAATGAAT | CATCGGAATC | ATCAGCACAC | 60   |
| TCTAGAACTA | CTCCATATGC | CGGGGTGGCC | GGGGGTCCCG | AGTAGTGCAC | GTGCCCATCG | 120  |
| GGAGACACAG | ATGATGGGTT | TGAAATGTCC | ATACGGGGCC | TGTGCACAAG | GGTCACGTCC | 180  |
| CCATCCCCAA | CACAAGGACC | TTTAGATAAC | CTCTCCCGGC | ATGTGCGCGT | ATCCGGGCAA | 240  |
| GCAAGCTGGT | GTTCTGGATT | CCAAACGTGC | CCAGCGGTAC | CCAAAATCGC | CAGGGCGTGT | 300  |
| TTTATTATTT | CCACAGGAAC | CGGTTTCTCT | AATTGCATCA | CCAGGGTATC | CAAAAGCCGG | 360  |
| GCTTCCACGT | TGATCCGGCT | TACCGACAGT | TCTTTCCAGG | GTTTCCTGGT | GGGGCCGGCC | 420  |
| AGCTGACTCA | AAAAGGTCAC | TGCTCTGCCC | CATGGGCGGG | TGGGTGACAG | TCCGCCATAC | 480  |
| TCTTCCAGGA | CACTGGCCAT | GCATGACTCC | AACCGTCTCA | CGTCCGAGGT | AATGTGCTCT | 540  |
| ATGAAGATGT | GGTAGAGCCA | GCAGACGTTT | AAACACGATG | AAATCAAGCT | AAGCTCCCGC | 600  |
| CGGAACCTCA | CATCCACAAA | GGGGTATTGG | TCCGGTGTCT | GTATTAGGTC | TGGAATAGAA | 660  |
| AAATCAGAAA | AAGACACTGA | CCCACCAAGG | AGAACCTGGC | GTCTTGCAAA | GTTGATGAGC | 720  |
| CCCSCAGAAA | GAATGTGTCT | CCCGTGGGAC | AAAGAGCTTG | GGGGGGCAGA | GATGGCGCTA | 780  |
| CAGTGGGTGA | TTTCTTCTAC | CACGGTCTAT | CATTGGTGGC | ACCCACAGGC | CTGTTCCAGT | 840  |
| ATCAGCATAA | ATCTATCTTT | GCAGTCATCC | CAGATCAAAG | TCATGTGAGA | TGCTGTTGCC | 900  |
| TGGCATTTTG | CCCGCATGTA | CATTTCCTGT | CCCACATATT | TTAACATCTG | TAATACTGGA | 960  |
| AGTAGATTCA | GTCTGGTGTG | GAGCCCCCCC | GGGGAAGCCA | GCCTATGCTT | CAGGACCAAC | 1020 |
| AGGGACGCTA | AGAACCCCGG | GTGTCCGGCG | TCCGGAAACA | GACCTCTGAG | AATACGCTCG | 1080 |

|  |      |
|--|------|
| GTCTTGACGA AACCCGATGT GGTACCGAAT GCCACAATCT GTGCCCTCCA GCTCTCACAA  | 1140 |
| TTTTCATCTC CAATACCCGG AATTGGGATA CACACCTCCA TGTTCACTCA CATGTACGCT  | 1200 |
| AGGGTCTCCC CACCCAACCC CCATAGGACC CAGCTACAGC TTATCCTCCA CTAATAACCA  | 1260 |
| GGCAGCTACC GCGGACTCAT TAAGCCCCGC CCAGAAACCA GTAGCTGGGT GGCATGACA   | 1320 |
| CGTCCCCCTT AAAAAGTCAA CCTTACTCCG CAAGGGGTAG TCTGTTGTGA GAATACTGTC  | 1380 |
| CAGGCAGCCA CAAAATGGC GCAAGATGAC AAGGTAAAGA TCGACCTTTT TATTGTATAC   | 1440 |
| TGAACAATGC GTGTTTACAA TGGTGTAGGT GGGAGCAGAG TTCGCCAAGC TCTACGTCCG  | 1500 |
| AACAGTCGGG TGTCAGGGCT CTTATTAAGT GTTCGGTGTG CTTGACCAA GCGGCGGAAC   | 1560 |
| CTAGGTTGGG TCTGTACAGG TCGTACCAGG CAAAAAAGGA TCGGGCGGTG CTTTTCAGGA  | 1620 |
| GAGTTAGGGA CGTGCTGATT ATGTGGACAA GCTTCTGCTC GTAAATGCAC CGCTGCTACA  | 1680 |
| TCTGAACGAC AGCTGTCCAA AAAAAACAA GGTTCAGCTG CACGTTAAAA TCTGTATCCT   | 1740 |
| GAAGTCTCT GTAAATGACA GTTCTACCA AGAAAACTT TTTTACCAGC CTGGCCATCC     | 1800 |
| ACTGAAAGGA GGGAGCACAC GTCCCGTTGT GCGTTGTTAG GATATCCCTA ACTTCGGAGC  | 1860 |
| GGAGACGGCC GSACGCTCCC ACAAATGGG AGAGGCACCA CTCTGTGCAG TCCGCGGTCT   | 1920 |
| GGGTTTCTGA TTCCAGGGGC GCGTGTGGG GGTATTGGAG AGTCAAAAT CTGGGCACTC    | 1980 |
| CCTTAATGAG CTCTCTCTCA AAACCTATGC AGCCAGGTC CACTAGTGGC ACCATGCGGT   | 2040 |
| TAATAACACC CTTATCTTG TCGTTGCCAA GTTGTACAA CTGCTGCAGG GAATAAGCCA    | 2100 |
| AATTCGCCCT AGCCGCGGGA ACCAGGTACG GCTCGCTTTG TCGGTGCTGG ACCAATATCT  | 2160 |
| GAATGGTCTT TGCAAGGTAT AGGGTCTTCT CAACGTTTAG AGCGGGTACG TGGCACTCTG  | 2220 |
| GATTGAGGCT GCGGACGGAC AGGGTATCTA ACTCCTGAAG TATCTGATCC CAGGACGGGT  | 2280 |
| AATGATACCT AAACAGATGG TTGAACAGST GATCTTTAAG GGGCCTTCTC GATGTCACTG  | 2340 |
| TAAAACTAT GACACGCCAC TCTCTCCTTA GGGTAAGAAG CTTCCGGCGGT CTTGTGTGGA  | 2400 |
| AAGCTTCGTC GGCCTCTGGG ACGAACTGAA GGCCCAACTC TACCAGTGTG TGCTCCTTAT  | 2460 |
| AAATGACGCA TACGAACCA TCTACGATCC CAGTACCTA AATAGASTGG TGGAAATGCT    | 2520 |
| GTGCATTCCG ATTATGAAG AATGTTCCAA GCTTGGTGGC CTATGTGGTC TGTTTACAGA   | 2580 |
| CATTAACATG TTTAACCTTT TCTGCTTTT TCGTGCCTCT CGAATGAGGA CCAAAGGCGC   | 2640 |
| GGCCGGGTAC AACGTGCCAT GCGCAGAGGC ATCCCAAGGC ATTATTCGGA TCTTCACGGA  | 2700 |
| GAGGATCTTA TTCTGCACAG AAAAGGCATT TCTGACAGCC GCATGCACCG GGGTGAGCCT  | 2760 |
| GCCTCCAGCC ATATGTAAGC TACTACACGA AATATACACT GAAATGAAGG CCAAATGCTT  | 2820 |
| GGGGGCTTG AGGCGACTCG TCTGCAATCG GAGGCCCATT ATGATATTAA CTTCTTCCCT   | 2880 |
| ACTGAAGCTC TACAACAGT ACGATACCGC CCGGCTGCTC TCTGAGCACT CCAGGGCCTT   | 2940 |
| CTGCTTTTTC GTTTTCCAAC CGGTCTACCT TCCGAGGATT ATGGGCGCCGC TGGAGATCAT | 3000 |
| GACCAAGGCT CAGCTCGCCC CTGAAAACTT TTACAGCATC ACCGTTCTTG CTGAGAAACG  | 3060 |
| CCGGCCAATT ACCACCGGCA AGGTCACTGG ACTGTCTTAT CCAGGAAGCG GTCTCATGCC  | 3120 |



|             |            |            |             |            |             |      |
|-------------|------------|------------|-------------|------------|-------------|------|
| AGAATCTTTA  | ATTTTGCCAA | TCCTGGAGCC | AGGACTGTTG  | CCGGCTTCCA | TGGTAGACCT  | 3180 |
| CAGCGATGTG  | CTGGCAAAAC | CCGCCGTTAT | TCTGAGCGCC  | CCTGCCCTGA | GCCAGTTTGT  | 3240 |
| CATTAGCAAA  | CCCCATCCCA | ACATGCCGCA | CACCGTCAGC  | ATCATCCCTT | TTAACCCTATC | 3300 |
| GGGTACAGAC  | CCGGCGTTTA | TTAGTACGTG | GCAGGCCGCG  | TCACAGAATA | TGGTGTACAA  | 3360 |
| CACATCCACC  | GCGCCCTTAA | AACCGGCCAC | CGGTAGTTCA  | CAGACGGTGT | CAGTCAAGGC  | 3420 |
| GGTTGCTCAA  | GGGGCCGTGA | TTACTGCGAC | AACGGTGCCG  | CAGGCAATGC | CAGCGCGGGG  | 3480 |
| TACCGGAGGG  | GAGTTGCCTG | TAATGTCAGC | GTCCACTCCT  | GCAAGAGATC | AGGTGCGCTGC | 3540 |
| ATGTTTTGTG  | GCAGAGAACA | CCGGAGATTC | TCCCGACAAC  | CCGAGCTCTT | TCCTGACGTC  | 3600 |
| ATGTCACCTT  | TCCGATCCGA | ACACGGTTAT | AGTGGCCCCAG | CAATTTCAAC | CACCGCAATG  | 3660 |
| CGTTACGTTG  | TTGCAGGTTA | CCTGTGCCCC | CTCTTCGACA  | CCACCCCCCG | ATTCAACAGT  | 3720 |
| CCGGGCCCCG  | GTGGTGCACT | TGCCAACAGT | AGTCCCTCTG  | CCGGCCAGCG | CGTTCCCTCCC | 3780 |
| GGCGTCTGCC  | CAACCAGAAG | CCTCGGGCGA | AGAGCTTCCG  | GGCGGTCATG | ACGGAGACCA  | 3840 |
| AGGTGTGCGG  | TGTAGAGATT | CAACGGCGGC | GGCTACGGCG  | GCAGAGGCGA | CAACACCCAA  | 3900 |
| ACGAAAGCAG  | AGAAGCAAAG | AGAGGAGCTC | AAAGAAGCGT  | AAGGCTTTGA | CGGTGCCAGA  | 3960 |
| AGCCGACACC  | ACGCCATCGA | CCACGACACC | TGGTACCTCT  | TTGGGATCAA | TTACCACCCC  | 4020 |
| CCAGGATGTG  | CACGCCACGG | ATGTCGCCAC | GTCTGAGGGA  | CCATCGGAGG | CACAACCCCC  | 4080 |
| GCTACTGTGG  | TTACCCCCCG | CACTGGACGT | AGATCAGAST  | CTATTGCCCC | TGTTAGACGA  | 4140 |
| AGCGGGCCCT  | GAAACATGGG | ATGTCGGGTC | GCCTCTCTCC  | CCCACTGACG | ACGCGCTGTT  | 4200 |
| GTCCASTATT  | CTGCAAGGAC | TGTACCAGCT | GGACACGCCA  | CCGCCTCTGC | GGTCACTCTC  | 4260 |
| CCCCGCTTCC  | TTGCGCCCGG | AGTCTCCGGC | GGATATACCG  | TCACCTTCTG | GTGGAGAGTA  | 4320 |
| TACGCAACTG  | CAACCGGTCA | GGGCGACCTC | GGCGACGCCC  | GCTAACGAGG | TACAGGAGTC  | 4380 |
| CGGCACACTG  | TACCAGCTGC | ACCAATGGCG | TAATTACTTC  | CGAGACTGAA | GTGTTCCGAA  | 4440 |
| GGGCGTCTGT  | GCCTGCGTTA | ACTTCCGAGG | CAGTTTATTT  | TTAACAGTTT | GGTGCAGAGT  | 4500 |
| GGAGTTAACC  | TACAGATTCT | ACTTAAATA  | GCTCATTTTC  | TCACGAATCT | GGTTGATTGT  | 4560 |
| GACTATTTGT  | GAAACAATAA | TGATTAAAGG | GGGTGGTATT  | TCCTCCGTTG | TCGACTATAA  | 4620 |
| CCTGGCGTGT  | AAACGTGTAA | CCCTGCCAAA | TGCCCAGAAAT | GAAGSACATA | CCTACTAAGA  | 4680 |
| GTTCCCCGGG  | AACGGACAAT | TCTGAGAAAG | ATGAAGCTGT  | CATTGAGGAA | GATCTAAGCC  | 4740 |
| TCAACGGGCA  | ACCATTTTTT | ACGGACAATA | CTGACGGTGG  | GGAAAACGAA | GTCTCTTGGA  | 4800 |
| CAAGCTCCCT  | GTTGTCAACC | TACGTAGGTT | GCCAGCCCCC  | GGCCATACCG | GTCTGTGAAA  | 4860 |
| CGGTCAATTGA | CCTTACAGCG | CCTTCCCAAA | GTGGCGCGCC  | CGGTGACGAA | CATCTGCCAT  | 4920 |
| GCTCACTGAA  | TGCAGAAACT | AAATTCCACA | TCCCCGATCC  | TTCTTGGACG | CTCTCTCACA  | 4980 |
| CACCACCAAG  | AGGACCACAC | ATTTCCCAAC | AGCTTCCAAC  | TGCGAGATCC | AAGAGGCGAC  | 5040 |
| TACATAGAAA  | GTTTGAAGAG | GAACGCTTAT | GCACTAAGGC  | CAACAGGGC  | GCAGGTCCGC  | 5100 |
| CCGTGCGCTG  | GTCTGTAGTT | AAGGTAGGGA | ACATCACCCC  | CCATTATGGG | GAAGAACTGA  | 5160 |

|             |            |            |            |            |            |      |
|-------------|------------|------------|------------|------------|------------|------|
| CAAGGGGTGA  | CGCCGTCCCA | GCCGCCCTA  | TAACACCCCC | CTCCCCGCGC | GTTCAACGCC | 5220 |
| CAGCACAGCC  | CACACATGTC | CTGTTTTCTC | CTGTTTTTGT | CTGTTTAAAG | GCCGAAGTAT | 5280 |
| GTGATCAGTC  | ACATTCTCCC | ACGCGAAAGC | AAGGCAGATA | CGGCCGCGTC | TCATCGAAAG | 5340 |
| CATACACAAG  | ACAGGTGCAG | CAGGTATAGA | CGGGAACAG  | GTGTCTATCT | TGGCCGCGTC | 5400 |
| GTTACTCAA   | TGGGAACAAT | GGCGCCACCT | TGCTGTCTTT | GTAGGCATTA | GAAGAAAAGG | 5460 |
| ATGCACAAC   | ATGTTTCCTA | GCGGCGAGAT | TGGAGGCACA | TAAGGAACAG | ATTATTTTCC | 5520 |
| TTCGCGACAT  | GCTGATGCGA | ATGTGCCAGC | AGCCAGCGTC | GCCAACGGAC | GCGCCACTCC | 5580 |
| CACCATGTTG  | AAGCTTGCTT | GTGCCGTCTG | CGGGAGAAC  | CATGCCAGAC | TTTGTGTGCT | 5640 |
| AAGAAAGGAAT | TGTTATCCGG | CAGCAATATT | AAAGGGACCC | AAGTTAATCC | CTTAATCCTC | 5700 |
| TGGGATTAA   | AACCATGAGT | TCCACACAGA | TTGCGACAGA | AATCCCTGTG | GCGCTCCTAA | 5760 |
| TCCTATGCC   | TTGTCTGGTG | GCGTGCCATG | CCAATTGTCC | CACGTATCGT | TCCGATTTGG | 5820 |
| GATTCTGGCA  | AGAGGGTTGG | AGTGGACAGG | TTTATCAGGA | CTGGCTAGGC | AGGATGAAC  | 5880 |
| GTTCCTACGA  | GAATATGACG | GCCCTAGAGG | CCGTCTCCCT | AAACGGGACC | AGACTAGCAG | 5940 |
| CTGGATCTCC  | GTCGAGTGAG | TATCCAAATG | TCTCCGTATC | TGTTGAAGAT | ACGTCTGCC  | 6000 |
| CTGGGTCTGG  | AGAAGATGCA | ATAGATGAAT | CGGGGTCCGG | GGAGGAAGAG | CGTCCCCTGA | 6060 |
| CCTCCACCGT  | GACTTTTATG | ACACAAAGCG | TCCAGGCCAC | CACAGAACTG | ACCGATCCCT | 6120 |
| TAATATCAGC  | CTTTTCAGGT | GTATTACACG | TTTCAACTGT | AATCCCTCGC | AATTGGGTAA | 6180 |
| ACCGTCGGTG  | TGTAGGGATA | AAGCGTAACC | TTACGTTCTG | TCTCATCTAC | AGGATCATAT | 6240 |
| TCATCTGGGG  | AACCATCCAG | GACCACGCGA | ATTCCCGTAT | CACCGGTGCG | AGAAAACGGC | 6300 |
| AGAAATASTG  | GTGCTAGTAA | CCGTGTGCCA | TTTTCTGCCA | CCACTACAAC | GACTAGAGGA | 6360 |
| AGAGACCGCG  | ACTACAATGC | AGAAATACGG | ACCCATCTTT | ACATACTATG | GGCTGTGGGT | 6420 |
| TTATTGCTGG  | GACTTGCTCT | TATACTTTAC | CTGTGCGTTC | CACGATGCCG | GCGTAAGAAA | 6480 |
| CCCTACATAG  | TGTAAACAAA | AACCATAAAA | GTAAATAAAC | GTGTTTATTC | TTACATGAT  | 6540 |
| AAAGAGTGGT  | ACTCTTTACT | GGTTTGGGGG | TTGGGTTGTG | GCGTGGTGSC | TGGTCCGGCG | 6600 |
| TTCAATCATC  | AACCCCGCGC | CGTGTGTGCG | AGGCTCCTCT | TGCTCGCCTG | TTATTGGCAC | 6660 |
| CAGGAGGCGG  | TTTAGCGGTG | CCCCCGTCTG | ACATGCAGAC | GTCGATTCTA | AGCGAAAGTC | 6720 |
| CCTTCAGGGC  | ATCGTCCACT | TGTTTTTGTG | TTACAACCTT | GCTGAATATT | GTCCTGACCC | 6780 |
| TGGCTTCGAT  | TTTCTTAGCG | GCCGCGCGAC | TCAGTGCAAC | CACAGTAGCG | GTAAGCTGCG | 6840 |
| CTTCTTTCTC  | GGTGGCCGTC | AGAGGCCGAT | CTCTCGGATC | GSCAGTGGAT | CCAGTSCCTT | 6900 |
| TCCGAAGCTC  | CCGATTCTCC | ACAGTCAATT | GGCTTATCTT | TSCGSTTAGG | TCTTCTATCG | 6960 |
| TAAGGTCTCT  | TTTGGGTCTG | CCCCGCGCGG | CGGCCATGTC | AGGTACGCGT | AGATGTACGT | 7020 |
| GTTGGTGATG  | CTCACAACAA | AAGCCCAAT  | CCCTCCTTTA | TACCCAGCTT | TAAATACCTT | 7080 |
| ATTGAAAAAC  | CATAGCTTTG | GTCAGCGCTT | GTGCGAGTAA | TCACATGCCA | GTCTATGCAT | 7140 |
| GGACCACTCT  | GTCCACAAC  | TTGAAAAAC  | AAAGATATAC | CAGATAGAAA | AATGTGGCCA | 7200 |

|            |             |            |            |            |             |      |
|------------|-------------|------------|------------|------------|-------------|------|
| CGACGACTAG | TAACGCGTTA  | ATCAAGGCCC | AGACGCTAGA | AAAGCTAGAA | AGGGAGGGGC  | 7260 |
| TAAAACTATC | CGCGGAACAA  | GCAACGTCAT | AGAATCCTGG | G3TAGTGAAT | GATGTGGGAC  | 7320 |
| CGGGCGAAGG | CCTGGCGCTG  | AGCCCAGCCG | TACTGGGACT | AGAACGCTCT | GTAATGATG   | 7380 |
| CGACACCTGT | CGAGTTGGCC  | GTAACCCAGC | AGTGACCTAG | TATCGAGGCC | ACAAATAAAG  | 7440 |
| CCAGGGCCAC | CGTGGACGCT  | GTCATTATGA | ACAACCGCCG | AGGCTCCAAG | CCGTCTATCC  | 7500 |
| AACGTTCCGC | GTTGCGCTCT  | TATATACACT | CTGCAATGCA | GTCGACTCT  | GCCCCCTCTAC | 7560 |
| CCAGGGTGGA | ATATGTGTTT  | GAAACAAGCA | AATTTAGAA  | GACGTGAGA  | GCAATGAAG   | 7620 |
| CCAGACTCAG | ACTGACAAAT  | GAGTGTCGGA | TACTGGTGAG | ACCCACGAG  | CCGTTCATCA  | 7680 |
| TGCCCCCCGG | AATACACTTC  | ACGCGAACCC | CTAGCTGCGC | TTTCATCCTG | ACCGGAGAGA  | 7740 |
| CCGACAAGGA | TGTATTTTGC  | CACACGGGCC | TAATCGACGG | AGGCTACCCG | GGGGAGATAC  | 7800 |
| AGGTTATTTT | ACTCAACAAG  | AGGAAGTACC | CTGTGACGCT | GTATCGCGGG | GAGCTCAACA  | 7860 |
| TCTGCCTGTC | TGCTTTCAAT  | TACGTGCTAC | CTCCGTTGAG | GGACGTATCA | TTCTTAACCC  | 7920 |
| CCCCATATGA | TGCAAAACGAC | GCCGGATTTG | ACGTGATGGT | GATGCACTCT | ATGCTTATCC  | 7980 |
| CTCCTACTAC | TSACCAACCG  | TTCATGATAT | ATCTAGGAGT | GGAGACCCCA | GGCCCCCCTG  | 8040 |
| AACCCACGCT | GGCTCTAGCA  | TTGGGGCGAT | CCGCTCTAGC | ATCTAGGGGT | ATAGTTATAG  | 8100 |
| ACGTTAGTGA | GTGGGGACCG  | CGAGGATTGC | AGCTGAAGTT | TTATAACTAC | TGGGGGCAGC  | 8160 |
| CGTGGCTGGC | GCAGCCCGGT  | AGCCGCATAT | GCCAGATTGT | GTTTGTGGAA | CGCAGACACA  | 8220 |
| TCCTCAAGGG | CTTCAAAAAG  | TGCTTGCGCC | ATAGGAAGCT | AGCTCCTGGC | GTCCCTTTCC  | 8280 |
| GGGAGGCTCG | AGTGCAATTT  | CGCGAGGATA | CAATAGCGT  | CCGAAACAT  | ACCCACGAAG  | 8340 |
| ACAACCCCGT | CCAGGAACCC  | AACGTAGCCA | CCGCTTCCGC | TGACATTGCT | GGAAACCAAG  | 8400 |
| GGCTGGGGTC | GTGTGGGTTT  | TAGAGCCGCT | GCCAAATGCG | GCCAGTTTAT | TAGGGCGATT  | 8460 |
| CGATCCCGCA | ACCCACAGCA  | TCCCCCAAAT | AAAAAAACGA | GTGTACACAG | CCAATGTTTT  | 8520 |
| TATTATTGTT | CGATTCAATTA | CTGGTACCAG | AGAATAAAGC | CAACCTATGT | CGAACCTATC  | 8580 |
| GCGCTTTCTG | TGCTCTCTTC  | CAGGGTTGAC | GAAGGCCCGG | GAGGGATTGA | CGAATGCATC  | 8640 |
| GCGGAACCG  | ACGGGTCTTC  | GGTGGGTGGC | TTGGGTAAAG | TTGCCTCCGG | CTGGCCTGTA  | 8700 |
| ACGGCTAGGG | TGAGAGGCCA  | TACAGAAGTG | GGTTCCGACA | AGGAGTGGCT | GATCTCAGAG  | 8760 |
| GCCCATATTA | CCGAGTCGTC  | TGACGCCATA | GCASTCGCCA | GTTTTTCCAT | CTCCATGAGC  | 8820 |
| GAAACGCATT | CCCCGGCCCT  | TTTGTTTAAG | AGGGACTGGA | GCGCACTGTC | GTCCACGGTA  | 8880 |
| ATCTGCGCGA | CCGCCAAGGC  | CAGCATTGTG | TTCCACACGA | CCTTCTGAAT | AGACTGCAGT  | 8940 |
| TTTTTCACCT | GGGTTTTTAC  | GGTCTCCTGG | CAGCCCGCCG | GAATTTTAGC | CACGTCAAAA  | 9000 |
| CGCTTCAGGT | AGTCTGTGAT  | CTTGTTTGAC | TGTACAGCCA | GAAGGTAGGT | CTGGTGCAGC  | 9060 |
| GCCGTCTGTC | CAAGGTTTGA  | CTGGACAACG | TCACCCAGAC | ACACTCCGGG | GGGGAGGCCC  | 9120 |
| AAATCTATCT | CTTGCCGCCA  | GCCTCTGGA  | CAGCCTTCCA | GAGGGTCACC | GAGGCGCTTG  | 9180 |
| TAAGCGTGGT | TGCCCGCTCC  | AAAAAGGTTT | ATACCGCAAC | ACGTCCAGGT | GTACCATGGA  | 9240 |

|             |             |            |             |             |            |       |
|-------------|-------------|------------|-------------|-------------|------------|-------|
| GACGACATAC  | CGCCGCGAGG  | CGCTGACAST | AAGGGTTATT  | TTTTGTACGA  | GTGGCGACAG | 9300  |
| CGCCGAGACG  | ATCGCCGACG  | TCCTTACGGG | GGCCCCAAGG  | TCAGCGTCCT  | TCTTTTCTGT | 9360  |
| ACTCCACGAC  | CTTTTTTTATT | CCCAGATACT | CGCCCCCAGG  | GTAACCCCTAA | AATTGTGCTT | 9420  |
| CCCCGACCGG  | CGTCTTGCCA  | ACGGCACAAG | GTGTTGCGCC  | GTGTTGGTCC  | TACCTACTGA | 9480  |
| CGCATCAGTG  | GCCTCGGGGT  | TCCTTGCGGG | CCGGCCACTG  | GAGGCGTCCG  | ACATTAAATA | 9540  |
| TATGCTGCTC  | AGCGACCAGA  | CCGCGGGGTT | GTTCAAGCCG  | CTGTTGGAGA  | TAATCGGTGG | 9600  |
| CGCGCGCGCA  | CCACCAAAATC | AGGACGCGTG | CACTTTCCAG  | AGCCAGGTGG  | CCTGGCTCAG | 9660  |
| AACGAATTTT  | GTTACCGCAT  | TGAGAAAAT  | TTACAAGATG  | ACTCCCTCAC  | CCTACTGGAT | 9720  |
| GCTGTCTGCA  | TTTGGCGCTC  | AGGAAGCCCA | GTTCTGTCCTG | ACCAGCTCAT  | TCTATTTTTT | 9780  |
| TGAACACACT  | GTGGTCTGTA  | CCACAGAGAC | AGTTTCTCAC  | CTGTCTAGAC  | TGTTTTCGCC | 9840  |
| TCAACAGGGA  | CAGACGCTGG  | TTTCCGTTAC | CAGCCACGAG  | GAGCTGGGGC  | AGCTATACGG | 9900  |
| CACTTCCCTT  | TTGAGGCGGC  | GCGTCCCCGC | GTTCTGTCCT  | TATGTAAAAG  | AGAAATTAGC | 9960  |
| GAGAGACAGT  | CTGGAGACGG  | AGGCCATCGA | CCGCACCATA  | GACCAGATCA  | GGGGCAAACT | 10020 |
| CATGCTGTCT  | AACCAGGACC  | TGGTCCATTT | CATATATATC  | TCCTTTTATC  | AGTGCCTCAA | 10080 |
| CAACCGGGCG  | TTCTTGCGCT  | ACTCTAGACA | GACGTCTCTT  | TCAAGTGCTC  | TAAGGGAGCT | 10140 |
| GGGGGAAGAC  | CCTCAATTGT  | GTGGCGCCCT | ACACGGGGAG  | TTTCTGTACC  | ACGTCCASTC | 10200 |
| CTACTACCAC  | AAAAAAACCT  | ACCTATCCAC | TTACATAGAC  | ATTGGGTACG  | TGGGTGGCGT | 10260 |
| ATTACCAGAC  | GGCTATTTTG  | GCGGGAGTCT | TGTAGGCGAG  | CGGTGCGTTT  | ATTGGTGGCG | 10320 |
| GCACTCAAAG  | GACACGGCCA  | GCCTGTTGGC | CACCATTAGC  | CAACAGGTGC  | CGCACTTGAG | 10380 |
| GTTGCAAAAC  | GAGTTGCTG   | GCATGCTAGA | CGTGGCCGCA  | CTGCGAGGTT  | CCGATGACGG | 10440 |
| TCAGTTTAAA  | GAGGGCCTTT  | TCTCCACAG  | TCAAGCCCTA  | CCCTGTACA   | GGTGGGASTT | 10500 |
| TCTGGGCAAG  | CAGTTTTTCA  | CAATGCTTCA | GGAAGACGGC  | CTAGAGCGAT  | ACTGGGAGCA | 10560 |
| AAGTGTGATA  | TTTCCAGGCG  | ACCAGGACTG | GGATATGTTA  | TCTGACAAAG  | ACCTCACCTA | 10620 |
| CCGAATTTTT  | TACCATGACC  | TCAGCCTATC | GCTGCCAACA  | CTGAAGGAAC  | AGCTCCTTGT | 10680 |
| TTCAAGACAC  | GAATACTTCA  | ACCTCGCTTT | GCCAGTGTAT  | AGATGGGTAT  | TAGACTTTGA | 10740 |
| CCTGCCCCGT  | TGCCCCGACA  | TTGACAGGAC | ATTGAGGAG   | GTGCACTCTC  | TCTGTTGTTT | 10800 |
| CCTGCGTGAG  | GCCATACTCG  | ACATCATTTA | ACTCCTTGGA  | CCAGTGGATC  | CTCGAACACA | 10860 |
| CCCAATATAT  | TTTTTCAAA   | CAGCTGTCTC | ACCGGACGAG  | TGGCGGGGCG  | AAGACGTCTC | 10920 |
| CAGCACCAGC  | TTCTGTGGGT  | GTCATGACAA | ACTGGGTATG  | CCTATTATCG  | TCTCGTTTCC | 10980 |
| AGAAGGASTA  | TGCTGTCTTG  | GGTGGGAGCC | CATGGTGGCA  | CTCACTGGCA  | TTCTAAACAG | 11040 |
| GACGATAAAG  | CTTGATCCGG  | AGCTGGTCCA | CAGATTCCCG  | TCAATACAAA  | AAAAGGGGGG | 11100 |
| CCCTTTTCGAC | TGTGGCATAT  | ACGGCCGAGG | ACGAAGCGTC  | CGGCTTCCCC  | ACTTTTACAA | 11160 |
| GGTGGGCTTA  | GTGGGGGAAC  | TCTSCCGCCT | ACTGAAGATA  | CTAGTCTGTC  | ACCCCGCCCC | 11220 |
| CAACGGCAAG  | GCGCACTACG  | TGCGGCGCGC | CTTTACGCTT  | CGCGAACTGC  | TCCATCACTC | 11280 |

|            |            |            |             |            |            |       |
|------------|------------|------------|-------------|------------|------------|-------|
| CCCCGGCCAC | AGCGCCGGTC | ATGTGGGCGG | AATCATCTAT  | AGCATCATGG | ATCGCAATGA | 11340 |
| GAATTTTTTA | GAAAACAAGA | CCATTAGCTA | TCTGCCGGCC  | AAAATACCTC | ACATCTTTCA | 11400 |
| GCGGATAGAG | ACCCTATCCG | GTCGTTCAAT | AGAGGACTGG  | CTACACTCGG | CGSTTTGGGA | 11460 |
| TAAAGCATAC | GACACTATAT | GTAATTTTTT | CCCAGATGAA  | AAAGCACAAC | AGTTTTCTCA | 11520 |
| CGTTGCATTT | ACGCAACAAG | GGGAAAACAT | CATCCAGTTA  | AGACCCCGTC | AGGGAAGACA | 11580 |
| CTTCCTCTGC | ATCAACCATA | ATCATAAAAA | CAAGTCAAAA  | ACAGTCCGTG | TATTCCTTAC | 11640 |
| CCTTCATTCC | ATTAGGGTGA | GCGAAGTCAC | GGTAACACTT  | ATGAGTCAGT | GTTTTGCCAG | 11700 |
| CAAGTGTAAC | AATAATGTTT | CCACGGCCCA | TTTTTCGTTT  | GTGGTACCAG | TGGGACTGGC | 11760 |
| CAGTTAATCC | CACTATATAA | CCTGGCTGCC | AGGTTCCCAA  | AATAGCCCCG | GGCATACGGC | 11820 |
| TCACTTCCCC | CCACATTCCC | CCCGTGCACA | ATATAAGAAC  | CAAAGGACAT | GGTACAAGCA | 11880 |
| ATGATAGACA | TGGACATTAT | GAAGGGCATC | CTAGAGGGTA  | AGTCCTCGTC | TACAACAGAC | 11940 |
| TTTTCCCATT | TCTAACGTAT | CGTGCTATCT | TCTGCGCCCG  | GCGGACCATC | CCCCACCCCC | 12000 |
| TCATTTATCG | CSTTTGATAT | TACAGACTCT | GTGTCTCTCT  | CTGAGTTTGA | CGAATCGAGG | 12060 |
| GACGACGAGA | CGGACGCACC | GACACTGGAA | GACGAGCAAT  | TGTCCGAACC | CGCCGAGCCT | 12120 |
| CCGGCAGACG | AGCGCATCCG | TGGTACCCAG | TCGGCCCCAGG | GAATCCCACC | CCCCCTGGGC | 12180 |
| CGCATCCCAA | AAAAATCTCA | AGGTCGTTCT | CAACTGCGCA  | GTGAGATCCA | GTTTTGCTCC | 12240 |
| CCACTGTCTC | GACCCAGGTC | CCCCTCACCA | GTAAACAGGT  | ACGGTAAAAA | AATCAAGTTT | 12300 |
| GGAACCGCCG | GTCAAAACAC | ACGTCTCTCC | CCTGAAAAGC  | GTCTCGGCG  | CAGACCACGC | 12360 |
| GACCGCCTAC | AATACGGCAG | AACAACACGG | GGCGGACAGT  | GTGCGGCTGC | ACCGAAGCGA | 12420 |
| GCGACCCGCG | GTCCGCAGGT | CAATTGCCAG | CGGCAGGATG  | ACGACGTCAG | ACAGGSGTGT | 12480 |
| TCTGACGCGG | TAAAGAAACT | CAGACTCCCT | GCGAGCATGA  | TAATTGACGG | TGAGAGCCCC | 12540 |
| CGCTTCGACG | ACTCGATCAT | CCCCCGCCAC | CATGGCGCAT  | GTTTCAATGT | CTTCATTCCC | 12600 |
| GCCCCACCAT | CCCACGTCCC | GGAGGTGTTT | ACGGACAGGG  | ATATCACCGC | TCTCATAAGA | 12660 |
| GCAGGGGGCA | AAGACGACGA | ACTCATAAAC | AAAAAATCA   | GCGCAAAAAA | GATTGACCAC | 12720 |
| CTCCACAGAC | AGATGCTGTC | TTTTSTGACC | AGCCGCCATA  | ATCAAGCGTA | CTGGGTGAGT | 12780 |
| TGCCGTCGAG | AAACCGCAGC | CGCCGGAGGC | CTGCAAACGC  | TTGGGGCTTT | CGTGGAGGAA | 12840 |
| CAAATGACGT | GGGCCCAGAC | GGTTGTGCGC | CACGGGGGGT  | GGTTTGATGA | GAAGSACATA | 12900 |
| GATATAATTT | TGGACACCGC | AATATTTGTC | TGCAATGCGT  | TTGTTACCAG | ATTTAGATTG | 12960 |
| CTTCATCTTT | CCTGCGTTTT | TGACAAGCAG | AGCGAGCTAG  | CACTGATCAA | ACAGGTGGCA | 13020 |
| TATTTGGTAG | CGATGGGAAA | CCGCTTAGTA | GAGGCATGTA  | ACCTTCTTTG | CGAGSTCAAG | 13080 |
| CTTAACCTCA | GGGGAGGGCT | GCTCTTGGCC | TTTGTCTTAA  | CTATCCCAGG | CATGCAGAGT | 13140 |
| CGCAGAAGTA | TTTCTGCGCG | CGGACAGGAG | CTGTTTAGAA  | CACCTCTGGA | ATACTACAGG | 13200 |
| CCAGGGGATG | TGATGGGGCT | ACTAAACGTG | ATAGTAATGG  | AACATCACAG | CTTGTGCAGA | 13260 |
| AACAGTGAAT | GTGCAGCGGC | AACCCGGGCG | GCAATGGGGT  | CGGCCAAATT | TAACAAGGGT | 13320 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| TTATTCTTTT  | ATCCACTTTC  | TTAAGGATTG  | CCAAACCCCA  | TGGCAGAGTG  | TCTCCCGTAT  | 13380 |
| TCCATGTAAC  | TCACGTAGCC  | TTTCTCTAAT  | AAACAAGCTA  | CCTGCAAACT  | ATACACAAAT  | 13440 |
| GAAATGAGTC  | AGGCCTGGTC  | TCTTCTCTAC  | CGTGAATCGC  | ACCTTAAACA  | CAACACCAGA  | 13500 |
| CCGCCACCAG  | GTGGCACCCA  | ACATCCATTA  | TGGAAAAACC  | CCGCGCCACC  | TTCCGCCACC  | 13560 |
| TGGAGCCAAAC | AAACAAGACA  | CACCCGCCAA  | TGTTTTGGTC  | TCTTTATTGA  | TATGATATAC  | 13620 |
| TCCCTCCCAT  | AACAATACGG  | TGTAGGCATT  | TTGTATTATT  | TATTGCATGG  | CATCCCATAA  | 13680 |
| CGGCTTCGGC  | ATTATTTTGA  | GTACGACGCA  | GGCGTCTGAG  | AAATTACTGC  | ACCTCGCCGC  | 13740 |
| AAAGTCTCGC  | GGGGACGGGG  | CGTGGGGCTC  | TAACTTGCCA  | ACCGCCACCG  | GTTTTCCCAAG | 13800 |
| CCACAGCTTC  | ACCAAAGGAC  | ACGTCACGTG  | AGAGGGTGCT  | GGTAACGGTG  | AATTTGCCAA  | 13860 |
| CCCCACCAGA  | AATGTATTGG  | GSTTAAATAT  | CCTCGTCGGT  | TTTCCCTGGG  | GCAGCAAGAG  | 13920 |
| GGGGCCGGAG  | TCAGGCGGAA  | CGGTATTTCC  | AATAAAGTGC  | ACGGGCCCCGT | TATGATAACA  | 13980 |
| TACGCCAAAT  | ATGCCATTAC  | AAGAGCTAGT  | CAGCAGAATG  | CCTTTTGCAC  | ATGCGTCCAG  | 14040 |
| CGTATCGCAT  | AGCTCCCGGT  | TGGCTATCTC  | GCAGGCCAGG  | TTTGGCACAT  | TGGGTAGCCA  | 14100 |
| TACCTGGCCC  | GGAGACCCCA  | CTGCACAGTA  | ATGAACTGCG  | GGGTCCCTAC  | GCAAGGCCGA  | 14160 |
| TGAGATTGGA  | CAGCCCTGACT | GGCTTGCTGT  | CAGTAACTCA  | TGAACCTGTT  | CGCCATTATA  | 14220 |
| ATACATCCTG  | ATAAACAACC  | GACCCCACTC  | AATGACGGCC  | TCCTGACCTT  | CTGCCGTCTT  | 14280 |
| ACAAGATGGC  | ACGGGGCTTA  | CAATCTCGCC  | TGGCAAGCAC  | TGCCCCGGGG  | AAAAAATCC   | 14340 |
| CTCTTGCAAG  | AGACGTGCCA  | TATTGTAAAA  | ATCGTGGACG  | GCTCCGGCCA  | CGACTCCACA  | 14400 |
| TTCCACGCAAT | TGTTCTTCTT  | CCGGTTTACG  | TACTCTAAAG  | ACCAGAAAT   | GGTGTCCATC  | 14460 |
| CTGAGAAATG  | CCTTTGCCAA  | TCTCTTGTA   | ACCCCGCGTC  | CTGCGTAGCG  | CGGCAAGCAT  | 14520 |
| TGCGCTGCGC  | CCCGTGGTGC  | CTTTAAACGA  | GSCCTCCACC  | GGCATGTTAC  | CCCTTTGCGG  | 14580 |
| GATATACACA  | ACACCCCAAT  | CCCGCTCTCT  | GGGCCATTCA  | AAACAGGGGT  | CCCGAGGGGG  | 14640 |
| CGTAACTGCT  | ATACGGAAAG  | GGGTGGCGTC  | TTGCTCTTCC  | CACCTCTACTC | CGGGAAATTT  | 14700 |
| TCCACTGTTG  | ACTTGACATA  | CTATCCAAATC | CTTGATTGAC  | GCTTTCCCTT  | CACCTGACCC  | 14760 |
| GGTAGATATT  | CTTAGTTGTC  | GTGTCCGGCT  | CCACTCCGTT  | ATCGCAGCCA  | CCACAGCCTC  | 14820 |
| CCGTGTAATA  | TGCGCTCCGG  | CTGCAGAACT  | CCCGGTCCCG  | GAGGGTCCCT  | CTCCCGGTGA  | 14880 |
| CTCCGACCTG  | GATGGTTTAT  | CCCAAGGAGC  | CCCGGAGCCA  | GATGTTCCCG  | GTGACCTTTC  | 14940 |
| TGACAAACAA  | GGTTTTTTGG  | GTATCGCCCC  | AGCGCGCCCA  | AAAGGGTTCC  | GTCTTTGGCC  | 15000 |
| TGSSGCCATT  | GTCCCCCAAC  | CAGACTAGCT  | CGCGCCGCCA  | TGTCCAGTGG  | TAAGCACAGC  | 15060 |
| TATGCCGGGG  | AGCCACCGGC  | CATCAGATAT  | AGAGAGGGCA  | CAGGCTCTCT  | ATATATCAGC  | 15120 |
| GCTAGGTGGC  | TGACATATTA  | GTGGGCCTAG  | CCGCAGAAAT  | GCCTGGGTAG  | TCAAAAACCA  | 15180 |
| GCGTTTCTCA  | AATTAAACGA  | AACTACATTT  | TTCTATTTTA  | AGTACGGGAT  | ACAAAGCAGG  | 15240 |
| GTCTGAGGCA  | ATCTGGCGCC  | CTCCACCCCC  | ACCCACCATTA | CCCCAAAAAG  | ATATGTCAGA  | 15300 |
| AAGAGCACTC  | TACCTATTAA  | CTCGTGGAGA  | AACATCATAC  | AAAACTGTGA  | CATTATTTTT  | 15360 |

|             |             |             |            |            |             |       |
|-------------|-------------|-------------|------------|------------|-------------|-------|
| AATACTTTAA  | TTTGTGCAGG  | TTTCTTCACC  | CCACACCTGC | TTTTTGTCTG | GTACAAAAAA  | 15420 |
| CCACTGCAGG  | GTCCCGCCTA  | TAGCCAACTC  | CTAAGCGGGT | TTTTTGCTAA | AGCACTTTTT  | 15460 |
| TAGACTGTCC  | CAGAAACCAC  | ATAGCTTCCT  | TTTCACTCAT | TTGAAAAACA | GGCCCGCCCA  | 15540 |
| ACTGCCTGGA  | GAATTTTCCA  | CCCCCTCTAC  | CATTTCGCGC | CTTTACCGCT | GGTGCAGAAAT | 15600 |
| CTAGCCATCC  | TATCACCAGC  | GATCCGCTGG  | ACCAATATAC | CACGCCCACT | TTTCGTAATC  | 15660 |
| AGCAACCCCTC | TACGCCTACA  | CCCCATATGAC | TGAATATAAC | CCCCAACAG  | GCTATGAAAT  | 15720 |
| CATGAATGST  | AATGTCTCTG  | ACACCAATCT  | TCCGCGGGGT | GGCGGCAGTG | CGACGCAAGT  | 15780 |
| ATCCACAATA  | AATGGTGCAA  | TAATTGGCGA  | AATGTCTGTG | CTGGTTTATT | TGGACTACAA  | 15840 |
| GATTACATCC  | GGTTTTTATA  | TTCCACATATA | TGATCAATGT | AGACTATCCC | AAATGGAGCT  | 15900 |
| TATAAAAATT  | TTAACAGTCA  | AGGGTACATT  | TTGGAAATTT | TCTGTAGATG | CCGGGGATGC  | 15960 |
| GGCGAAAAAT  | ACCGTCCCGC  | ACGTCACTGG  | GTTGACGCTC | AGCGGTGTCT | GTGGGATTGC  | 16020 |
| GGCTGTGGTT  | GCCAGGTATC  | GCGCGGTGTT  | GAACAGCTGC | TGCGGAACTC | TGGGGCTAAA  | 16080 |
| GCTTGGGAGG  | ATGCTTTCAT  | AGCGGGAATT  | TGGATTACCA | AACCACCAGC | CTTCCACTTC  | 16140 |
| AGTGGCGTTT  | CTGGAGTATA  | TTCCAGACAT  | CGAGCAAAAT | ATTGGGAATC | CGTGGCCAAAG | 16200 |
| GCCTTCAAAA  | ACTCGSTTCA  | AAATCTCCAT  | TTCTCTGGGT | GAGGGGACTG | TAAGACGCGG  | 16260 |
| TATGCGAAGC  | AGTTCTGGTA  | CGAAACTCTG  | ACATAGGTGC | CCCAACGTAT | CCCCAACAGG  | 16320 |
| CCAGCTACAT  | AACATTGCCT  | CGCCCGCGTC  | ACCTTCGCGT | CTCAGAGTTC | CACGAAGGTT  | 16380 |
| CCCATACACA  | AAGATTTCCA  | CAACAAAAGA  | CACCCGCTGA | CTATCAGGGG | GATCAAAAAA  | 16440 |
| CATCTTTGAA  | GSTGGCTTTT  | CGGGACCGGA  | GTGGCTAACG | GGCGTACGCT | GGCCGTGCGG  | 16500 |
| GGACCTGGAC  | CTCGGGCGCC  | GGCTATCCGT  | GGCTGTCTG  | GTTGAGGAGC | TGGSTTCCTC  | 16560 |
| CTGCAGCTCA  | GACAAAATGT  | TACCCAACCC  | TTCTTCCCA  | GTACATATAT | CTCTCTCTTC  | 16620 |
| AAGGTTTGAG  | AGCGTAAGAG  | GGAGACCCAA  | AGGCGGCGGC | ACTAAAGATT | GTTCTGTGCT  | 16680 |
| ATAACCCCCC  | ACTGCATATC  | TATCTCCAGC  | ATATGTACTA | ACAAGTGGAA | CTCTGGGCGT  | 16740 |
| TTCCGCCACTA | CCCGGGGCACA | CACACTCCCG  | CGCTCCAGC  | TCTGTGGGTA | AATGCGAAAC  | 16800 |
| CTCGGGGTTT  | ACAGCGGGCT  | CGGTGCAGA   | ATAAAGCACC | GTAGSTTGGG | AAACGCGCGG  | 16860 |
| CCCACTGACA  | GSTAGGGGCG  | TGGATGCTAC  | AGTGGTAGAT | GGGSTATCGG | AATCCCCAGT  | 16920 |
| GAGSTCAATA  | ATCTCCACTT  | CGAGGGCACC  | AGAACTAGTT | GTCACGCGTC | TGTATCCAGT  | 16980 |
| CGCCATSTTG  | TCCCCCTGGC  | AGACCTACCG  | TATTCCAGAC | GAGGATGGGT | CTGTCTCGCTC | 17040 |
| TGCCACCTCT  | GGGCTGGGTG  | GTGCGCGCGC  | GGAGGGCGTG | GCCGACGCGC | CACCTTGCGT  | 17100 |
| GTGGGAAAGA  | CCCTGSTTTG  | GAGTGCCTCC  | ACTAGACCA  | GGAACTCAAA | GGGCTGTGCG  | 17160 |
| AACTTCCGSC  | ACCACGGCGT  | GACCAACTGG  | TGGGTGCCAA | ACAGGCGCGC | STATGGGTCC  | 17220 |
| CGTAGCTGGC  | GSTTCTGCCA  | ATGGACTCCA  | ATTGTAACAT | GATGSTTTTC | CATACCCGGG  | 17280 |
| CGCGGGGGCG  | CTGGGCGGTT  | GAGGTTCCGA  | GGGATACACC | CGTTCACCTC | CAGCACCCCTG | 17340 |
| AGGAGCCCCG  | CTTTCTGTAG  | ATGCCCCGCA  | AGCGCCTTCG | GCACCGGTTT | CCCGGCGGGG  | 17400 |

AAGCCACGGG CGAGCACATT GGCCGCTTTG GGGGAGCAAT CCTGTG33G CCAGAGGTGC 17460  
ACCTTGGGTG AACTCACCGA CAAATGTTCC CGCTTGGGGG TGCGGCGGAA TCCAACTGGG 17520  
GGCAGCAGGA TTCAGCTGGC TGCTAGGAAT CCCCCTATAT GTCCAAACGG GGGAAAGGGG 17580  
ATCAAAATGG CCGGTGGTTG GCGGATGCAC TTTCTCCGGG AGACCAGACG CGCCCTGAGG 17640  
CCACCATCCC GTGACAGGAA GATCTCCCCA TGGAAAACAC GCAGGTATCC ACGGGGACGT 17700  
AGATGGCAGC CTAGACCCAT CGCGCATGGG AGGGGCTAGT TGCCCCGTAT CCCCCGGCGT 17760  
CTGTGCGAGC CCGGAGACCC CTGACACAGT ACCGGCAAGC CGTGTTCGT GCTGCGGCTT 17820  
GGGCGGCGCC GTGCGCGGTA GGCCTGCACC AGATGAGTGA GGGTCTGAAG GGGCGGTGAG 17880  
CGTTGATGGA GCAGGCGGAT CTCGGGGAAC CCGCCACGTA AAGGACGAGG CCTGCGTAAC 17940  
TTGTGCGCTC CCAGAGGACC CCATACCTGA GGTAGATGCG CCTCATTCCT CTGSTATCCA 18000  
CACGGAGCAG GCAGCCTTCT GTTCAGTCGT TATATCGCCA ACATTGTAAT AGCGGTTCTGA 18060  
TTTCCGAGGG CGACCCCTCA GCCCCGATGG CGCCTTAGGG GGAGCAGGTG CTGCAGCCCC 18120  
TGCCCTCCTG TAGCTTTGTT CTCTAAGTAA AAGGCACGAG AGTTAACGTG GTTAGGGTAC 18180  
CTAAAGTATT TCCCGCCGAC ACCAACGCAT CAAACCTCAC ACCCCCTTCC CCGAGTTACA 18240  
TACCTAGTGT CACTGCGTCC CGTAGCCGTG GTTTGCATTG GGGGGGACAA CAGACACTGA 18300  
ATAAATCGCT GCAGTTTTTC AGGACCATAC GCGGCCCCAT AGCAATACGT ACAGTTTTTA 18360  
AACGGCGTTG GCACCAACTG CCATACTAGG TAGCTACCAC CAAATGTGTC GCTGTACCGT 18420  
AAATCGTTCC GCACGACGGC CCTCCTGGTT CCACGCAACA GTCTCCCAA ACGTCCATAC 18480  
ACCGTCTGTC CCACGACAGG CGATGGTCCG TAGACTCTAT CACACTCCTC ATCAAAATGCA 18540  
TGCTACACCG AATACCAGCC AGGCGGGATA TCGCTGCCCG CAGGCAGGGG CGCGGGGGCT 18600  
GCAAAAGAA GGTGTTCCT ATCAAACCG GAAAAATAGG GAACTTATT GTTTTCAAGG 18660  
GCATCAATAA TCCATAAGT GGGCCATTCT GAGCCACCGG CTTTAGGCAT GGTCCGACAC 18720  
AGAAACCGAT CCGCGTTCTG CTTTGAGGCA CAGTCCCGAC TGAGCCTTAT AGTCCCCCCC 18780  
TTCTTGCTAT GAAAAAACC CACGACCGTT ACGCAATTT GAGGAGCTAC TCACCTAAAA 18840  
GTAGCTCCTT TGACAAATGT CCGGTGTTTA TACCAATTGT TCACAAATGAC ATATTGTGCT 18900  
GGCGGAACA GGTGTCCCGA TGTATCCTCG GCAAGTAAGC ACCATTACCA TGTGCCATCA 18960  
TATTGTGTGG CACAAAAAA GCAACTTTTC ACGCACGCGC CATAAGACCC GAGCCAGTCC 19020  
CGCCCTCCAT CGCGCCTGCG AATTTTCCCA CCACCCATA TTGTGCGAGA TCTTTCTTAT 19080  
GTATATGTGG TTACAAACAC CACGCCCTTT AAGCTGTCTT CTCTCCCAAG GGGACTAGAT 19140  
TATAACAGTG ACATACGAAA CCGAGACGCT CTCAAATGCT TTCTATTTTA TTTATCGATT 19200  
CCGGGTTAAC ATAATCACAG GTAGCTATAA AATCCCCATC CTCTTGACCT GGTAAACCTG 19260  
GCTTGAGGTT TCTCTGTGTA TCAAAACAA CTGACCACAA CTGTACAGAG AAAAGTGGGT 19320  
GAAATGTAGT GTTTATTTTA TCTCACACT TTCCTTAAAC CACAGCCCGT CAAACACAG 19380  
GGACCTGTT GGTGACTAT TAGTCATCAC ATGTAAGTGA ACGCAATCTG ACCTTGATGA 19440



|             |            |             |            |            |             |       |
|-------------|------------|-------------|------------|------------|-------------|-------|
| CGAGGGGGGAC | CATATCGAAC | TGTTCTGCCG  | ACGTTGGGTC | ACCTCCGATG | AACACAGTTG  | 19500 |
| TTTTTTTAAT  | GTGCTCATGT | CCCTGSTATG  | GATATTGTGC | CACATTAAAA | ACATCCAGAA  | 19560 |
| CAGCCCTAGA  | TGACAGTCCG | CAGATCACAC  | CAAACTTCTT | TGGAGGATTA | TTTCCATGAT  | 19620 |
| ATAATACGGT  | AGACTTGCAC | AAATTCTTAA  | CATAAATGCC | AGATCGGAGA | GAAACTATCA  | 19680 |
| CAAGACCCGA  | AGCAAACGAG | CGCAGCACGG  | CGGCCAGCAG | GTTAACGTCT | CCTGGCCCTG  | 19740 |
| TGTTATTGTC  | GTCAGGTTTG | GGCAACAAAA  | CTCTTAACCC | TTTGGCGGAA | TGCAAGCAAG  | 19800 |
| AGTGGCTAAT  | GTCTGCCAGT | GGGTTCTGGG  | AACATAGAAT | AAACACCTTT | CGTTCCACTT  | 19860 |
| CCAAAGACAT  | TGCAGGGCGG | CCAAAATAAA  | ACACTTCCAC | ACCAAGCCTA | TCGGTTATCA  | 19920 |
| TTACTGGCGG  | CCGTGCCACT | CTATAATATG  | CGGATCTAAG | CTTCCTGTGG | CGAATGCGCC  | 19980 |
| TCGTGGTAGG  | CCTCTCGTGT | CTCCGTGGCC  | CATCATCCCA | TAAAAATTCC | CCAACAAC TG | 20040 |
| GCCGGCGTCT  | GGACGCCCGC | GGCAGTCCAG  | CACCATCATC | GACTTCTTCG | TCACTTATCT  | 20100 |
| CCAACACATA  | TTCCCTTGCT | ACATTCTGCG  | CCTCGAGTGC | CCCAGCTAAG | TACACATCCT  | 20160 |
| CTACACCCGC  | CCCGACAGCC | GAGGCGGCGA  | TTGAGCCCTC | TGTTACCACG | CCGCTTGCAT  | 20220 |
| CCGTGTGCGC  | TCCGGGCTGT | GATGTTGCGA  | TAACATCCTC | TGGGATGCCA | AGCAGATCAA  | 20280 |
| AGAGGTCTTC  | ATCGCACATC | GCCCTCATTG  | GCATGTCCAT | CTCCTGTCCC | ACGTGGTACA  | 20340 |
| TCAATGCACA  | TGCAGATTCT | TTATCAAGCA  | GTGTGAGGTC | ATCTTCAACG | TTGTCTGTGT  | 20400 |
| GCACCGTTGT  | TTCATCGGCC | GGGGGGGGCT  | GCGAGTGGCT | ATGACGCGTC | GAGGSTCTTT  | 20460 |
| CGTCTCCAGA  | GCCAGGAGAG | TGGGCATTGG  | CATCATCAAC | TGGCTGAACC | CCAGACGCAC  | 20520 |
| TATGGCGCGT  | CGATGGTCCC | TGCTCTCCAG  | AGTCCTCAGA | TTCCGCGCCC | GTCTGCGTGA  | 20580 |
| CCGGCACATC  | GCAAAAGGCT | GGGTGATCCT  | CCTCACTGGA | ATCCGAGTTT | TCACCCACAA  | 20640 |
| ATGGCCTACA  | GAATAAATAA | CAATATGTGC  | AACCGGACTA | GGGTGGCCAA | ACCATTGTGC  | 20700 |
| CCACCCCTCC  | CCACTCTTTT | CCCAGGGGAC  | ACATCTTACC | TTGGTCTTCT | CGGATGCTTC  | 20760 |
| TCCAGCCCTA  | CACTGTGTTG | ATACAAAATT  | TCCCATAGTG | ATGACCCACT | GTGTAGGTGA  | 20820 |
| GTCTTGGCAT  | GAACGCACCA | CCAGCATTCC  | TTTACCTCGG | CACACAGGAG | GCGCCACCTT  | 20880 |
| CTACAATTAA  | TTCCCTGTAC | GACCTCGTAC  | TCTTCACCTG | GCAAGCGTCT | AAGGCGCCGC  | 20940 |
| GACGTGGTAC  | ATATTTTCCC | AAAAGCCGTA  | ATCGGCGAGC | CCAGTAAATC | TCTGGGATGC  | 21000 |
| AGGCCCTTCC  | ATAGGCATTG | CCTCTTAAAA  | TCAATGAAAA | ACTGTAGGCT | ATCCAGAGGA  | 21060 |
| ATTAGTCTAT  | TACGGGCAGC | CGGAGCAAGA  | AATGTTCCAG | TAGATCTATC | TAGCCACTTG  | 21120 |
| ACCAAAGGAT  | ATTTATCAGA | GTCCAAAGCA  | CCTACAATAA | ACTCAGAAAT | CCAGGTAAGC  | 21180 |
| CTCGCTCCCG  | CCATGTTGAC | CTCTCAGAAAT | GGTCTGCCCT | CGAGCATTAC | CCCACCTCAA  | 21240 |
| CAGAAGTAAT  | CTACTACGCA | AACCACAACA  | TGCTTCCTGC | AGCTTTAACC | TTCAAGTCCG  | 21300 |
| GGTCAAAAAG  | CATTGCTCTG | ATTAGACACA  | TGTGTTTCTC | ACTATGAATC | GTGCTCTCCA  | 21360 |
| GCGCTGGCAA  | GAACATCTGG | GGTGTATGCTG | CCCCGGACCA | GCTTTGAAAC | AGGSTATTTG  | 21420 |
| ATGCATAATG  | AAGCCACAT  | GTTTGTCTTA  | CTTTACTAAC | CTCATTACCT | TGCATTGCAG  | 21480 |

GGGACACCCC CTTGCCTTGG CAGCTGAGTG AATCCCCAACC GCCTAGGAAA AAAATAACCA 21540  
 CTCAGACTTT ATTTTGCAGC CACACGGTGG CGCTAACCTT TAATGATGTC CCACTCAGTG 21600  
 AGTTTGGCCA CTCCCAAGCC CACATGGGGC TACTATAACA GGAACATAG AAGTTGCGGA 21660  
 TAGAGCCTGG TTTCTAACGG CAATGATATT TATAGTGCAA AACGGAGGGC GGTAAAGCAA 21720  
 AGGGAGGTAC CCGGACAGAG TGACAAGAAG ACTTGTCAA ATTTTAGTCT CTGTGCTAAA 21780  
 ATGGGGCAAG GTAAATGTGC AAAATGACTG GATAGTGATC CGAGTCATAT TCAGGCGAGC 21840  
 GCCGGCGGGC CAGAAACAGG GACGCGTACC GGGACCCCTC AGGTTCTCGA TTATGTCGCT 21900  
 CCACGTCAAA ACCTTGTGG ATCTCGTGGC GGTGGGACAG GGGCCTACAT TTGCTATTTC 21960  
 TTCTTCGCGA TGCATTTCOA ACAAGTATG CTGGGTATTC CAATAATCCC TTCAGAAAA 22020  
 TGCCCATGTT TGTACCGATG GCCACAACCT CCATGGAAAA CCTGTCCAGC GTCTGTTCOA 22080  
 AAGTTCGGTT TGGCTCCACA CTACAGTGGG CCGTTCTGGG AAGTAAGCAT TTATACGGGG 22140  
 GTACCGTCTG ACATATGTGT TCAGGGGAGG CCTCTGGGAC TTGGGAGCAA ATAACGATGC 22200  
 CCCCCGTAA ATCAAAGTGG GTCTTCACCT TTTCTCCGAA ATAATACACT TCCACCACTA 22260  
 GGGGCACAAG CTTGTACCCC ACTTTGTAAA TAGCCTGTTT CTTACTCAGG TATGCTGCCA 22320  
 CGGATTGGGT GCGGTTAAG ACCTTGGGCC TCATGTCGCT TCCATACCAG TAAAATGTCT 22380  
 GGTGAGCTTT CTCTTGTCTC TCGACGTCCC GGTGATCAGC ACACACGGGT GGAATACAA 22440  
 CAATAAATC ATCCACATTG TCGGAAGCTT GGAAGATGA ACCCATGACA GAGGCCCCAG 22500  
 GTGCCGAAC CTCAAGGGGA TGCGTGGCGG GAAGTACTGA GACACTCTCC GTGGACCCCT 22560  
 CCTCAGCTCC CTCCGACTGC ATCGGGCCCT GAGGGGTGCG AGTTTCACAC AGAAGTTCAC 22620  
 TCAGGTGCGC TAAGTCAGGA AGCTCCTGGC CTGAACCCAT GACAGAGGTC CCAGGTGCGC 22680  
 AACTCTCAAG GGGATGCGTG GCGGGAAGTA CTGAGACACT CTCGTTGGAC CCTCTCTCAC 22740  
 CTCCCTCCGA CTGCATCGGG CCTGAGGGC TCGCASTTTC ACACAGAAAT TCACCCAGGT 22800  
 CGCCTAAGTC AGGAAGCTCC TGGCCACAT CTGACAAGAG ATCTAACAAA CACCCCTCAA 22860  
 TGTGATCCAC CATCGGTAGG CAATCATCCA GCCCACTGAC ATGACTGGGG ACGGGGCTT 22920  
 CTGGGGAAAA TGGGGTTTGC GACTGTCCAG CAGGCGGGGC TAATAAGCCT TGTGTCTCAT 22980  
 GTGGAAAAAT AACAGGAGAA GGTAAACCCC CCGTTGGCAA ACATAGATCC GTGCGGTGT 23040  
 GCACGTGTAA TGGGCCCTGC ACCTGGCTCG TGGAGGGAGC CGGGGAATCC GGAGCTAATA 23100  
 AGCTCGATGA CTGACCAGAT GACCCAAACC CCGACGGTTC TGGCTCTTCA AAAAACAAAC 23160  
 TGTGCATATC CCTCCCTACA AAACCTGAG CCCCCACCCA AAGTTGTTTT TCGCTGTCA 23220  
 TCGATTCCGT ATCTTCGCTC TGTGACCGTG ATGAAGCTTC AGCTGCGGAG GATTTTGTG 23280  
 GCSTGGCGAC TGGCGCGGCC TGTTCCTG GGGCCTCCCT AAACAAAAT TAATTACACA 23340  
 AAGGTAAATC TGAGTGACAT CTCCAATTTT CCGTGATGCC CGGTGCACGT ACATCCCGCC 23400  
 GCCACACAAA CCGACCGCCC AGTACATCAA CCATCCTACC TCTGGGCTTT TTTTCTAAG 23460  
 CTCTTTCTAA GTGCCTTTTC TCTGTGTTTG TCATCATGGG GATAGATCCC AAACAATGCT 23520

|             |            |             |            |            |             |       |
|-------------|------------|-------------|------------|------------|-------------|-------|
| TTTAGCATGT  | TTTTCATGGC | TGGTTCTGTC  | GTCAAGTACA | CAAGACATCC | TTACATCCCC  | 23580 |
| TTGTATGGCC  | TAGGTGTGAT | AATCCAGCGG  | TTGAGTTTCA | TTTTTCCCTT | ATAGATGGTA  | 23640 |
| AAGGGCCCTCT | CCGTCTGGGC | TCGATTGGCG  | GTCTTAATA  | GCCGTCCAAA | GCAGCCCGAG  | 23700 |
| CCAGTCTCAG  | TCTCCGGGAT | TTCTGGCAGC  | CCGTGCCTAC | GTGCTCCTC  | CAAAAATGCC  | 23760 |
| TCATAGAAAT  | CATCGAAGCC | TTCTGGCATT  | CTCTCCCGCC | GGTTTCGACC | CGGCACGGTG  | 23820 |
| AATATTCTCT  | TTGTTCATC  | CAACCACCCT  | ACCCCCCAGA | AGCGTCCACT | GTCTAAAGCA  | 23880 |
| TCTATAATAA  | AGTCCGTGAG | CCATTCCGAC  | TCCGTGTAGC | GAGGCATCTT | TTTAGGCAAA  | 23940 |
| AGCCACGACA  | CAAAACACCT | TTTCCGTGGG  | CGACTTTTCT | GCCACAATA  | GCTGGACCCC  | 24000 |
| AACCCCACTG  | GCACGTAGAC | TCTGTGCCAT  | CTAACAACAA | AACTCAATAT | ATGCAGCTCA  | 24060 |
| ACACCGCCCC  | CCCCAGCCGG | TTGTGGGGCT  | GCGGAAACTT | GTGGTTAGAA | CTCACTACGG  | 24120 |
| AAAAGGGAAAC | CAATGCAGTT | GAACACTGCG  | CACACACCCA | TAACCCGGGA | CAGCACCCAG  | 24180 |
| GCACTGTCCA  | CCCTCTAATA | CAAGCGGCCCT | TTGGACGCGA | GGGAGGGGTG | TCATGGTCAA  | 24240 |
| CAAAACCAAG  | AAAACACATG | TATTATTCAA  | TTAGCCAACA | ACTTTATTTA | TTACCGACAG  | 24300 |
| GAGACATGAG  | ATACATAAAT | TTCCAACCGT  | GCATAGGGCC | AATACCATCT | GTGGAGCGTT  | 24360 |
| AAGTGCCTTG  | TGGAGTTTTT | GCCTAATTAG  | CTGAATCTCG | ACCCCCATTG | CGGCCAGCAT  | 24420 |
| GCTCACGAGG  | AATAGGCAGC | AGAGGCAGGA  | CCTAACTAGG | AGCATATCCG | GACCTGATCC  | 24480 |
| AAGTATGTTC  | ACCAAGGTGA | GCAACACTGC  | CGCCAAAGGC | AGGAGAACAA | ATAGCGCTCG  | 24540 |
| TCGGGAGGGG  | ACGGATACGC | CCACGCATGA  | CAGTAACCCA | ACATAAAATA | GCGTCATATA  | 24600 |
| CTTATCCAGG  | CCAATCAGGA | CCGGAGTCAG  | CAGGCCGATC | GAGGCCGTCT | ATATCAGGGT  | 24660 |
| GGCCAGCAAT  | AAGGTACAAA | ACACGACAAAC | CTCGCGCCTA | CAGTAGGCCC | AGGCTTGGA   | 24720 |
| CACTGAATAG  | GTGATGTACT | TCCCGGGCAT  | GATGAATATG | GCCCTCCTCC | TTTGCATTCC  | 24780 |
| GGCCCTGATG  | TACACATGCT | GTTCCAGGTG  | CCTAAATGCC | AAAAGTCCCC | CGACCAAGAA  | 24840 |
| GACAATGAAG  | GSCAGCCAGA | AAACGCCCGGA | CACAAAGACC | TTCTTAAACA | ACAGAAAGGA  | 24900 |
| GTACACCATA  | AATGCTCCGC | AGAAGCCCGAG | CTTATAGTAC | CTGTGTACTA | TTGGGGGGCG  | 24960 |
| CTGATACACC  | GCCGTTGCGG | TGGCTAGCGG  | ATAAGGTAAC | AGCAGTAAAC | AGTTAAGTAC  | 25020 |
| GCACAGACCC  | GGTATGAAGG | GCACACGAGA  | AAATGTAAAC | CCAGAAAAGG | CCGCGCAAAAC | 25080 |
| TACAGCAGCA  | AACACTGCTG | ACGCGCAGAT  | CCATTCCAGC | CTCCGGTCCA | GCTGTFTTTG  | 25140 |
| CGCCGCAGGG  | CACAGACACA | TGCATATCAG  | GGCCAAAGTC | GTGACTGGCA | GCGACCAGAA  | 25200 |
| AAACACGGCC  | GTGATCTCTG | TGGTAAAGAG  | TGTGAACGAG | TACAGGGCCT | TGAAGATAAA  | 25260 |
| ACACCACAGA  | AAGGGGGTCT | CCGCCAACGT  | CCCGCTCAGA | TAACTGAAGA | GCGACAGAGC  | 25320 |
| GCGCTCACTG  | TCCAGGGCGC | ACATGGTGTC  | AAATCAGGGG | GTTAAATGTG | GTTTTGGGCA  | 25380 |
| CCTTCCCAAG  | ATCCCTGGAC | TGGCTCGAGT  | CTGAGCGCCT | CTTGTGAGGC | CTCTTTGTGC  | 25440 |
| TGTCTTAAT   | TGGCGCCGCT | GGGGGGCAGC  | TGGTGACAGA | GGCAGCGTCC | TCAGAGGCCT  | 25500 |
| CCTCCAGCGG  | CCCAAAGGGA | CCAACCTGGT  | TGAGAGGGGG | AGAAATCCGA | GACTCCAAAT  | 25560 |

|            |            |            |            |            |             |       |
|------------|------------|------------|------------|------------|-------------|-------|
| CCGGCTGCCT | CCTGGAGTCC | GGTATAGAAT | CGGGAACCTT | TTGCGAAGAC | TGCGCTCCCT  | 25600 |
| CGGCAGACAC | AGATCGGTTT | ACCTCTAAAA | GTAGGACACT | TAACTTTACG | TCACCTGATT  | 25600 |
| GGCAGCCAST | GGGCACACCT | TCCACTTCTA | ATATTTCTGT | GGAGTGCCAA | ATCAGCCCGG  | 25740 |
| GGGTAAACCA | ACCCGGGACT | TTACACAGTC | TCAGGGCGGC | GATTAAGGAC | TCCAGGCTAA  | 25800 |
| CCCCGGTCAG | GGCGTCGGTG | TGCACCACGC | CCACATCCAC | CGACTTCTTC | CCCTTCAGAC  | 25860 |
| CATCCCAGCC | AGAAACGGGT | TTGGTTTCTG | GCTTGAAATC | AATGATCTTG | CTCAGGCCAC  | 25920 |
| CAAGAGAAAA | TGTCACGATC | GACAGCGTCT | CGGTGACAGA | CACAGTCACC | GTTTGGTCCCT | 25980 |
| CTTTTGTCTT | TTGCTGCCTT | AGCCACTTAA | GTAGGAATGC | ACCCGTTTTG | CCACAGAGGA  | 26040 |
| GAAGCCTGGT | GGTCTTACCA | CCGGCTTCCA | TCCGATCGTG | GAAAGGTAGG | ATACCCCTTT  | 26100 |
| GGTCCACCAC | GCTTTTGTGC | ACGGTGGAGG | TGAGGTTGTC | CCCGTAGGAA | ATGGTGGTCC  | 26160 |
| TGACGAAGTG | CGGTTGGGCC | CCCGTATCGC | ATGCTCTCCC | CTTTCGATAA | AAGGCTATGC  | 26220 |
| CAGCGTCGAG | TACATTGCGA | CCGAATAGCT | CACGCGTGTG | CGTGAAGCCG | CTACCGACGG  | 26280 |
| ACGTATTCCT | GAAGCTGAAG | CTAACGTCTC | CACTGCCTTC | CGTGTGTCCC | ACCAGGGGCG  | 26340 |
| TAAGGGCATT | CTTTATTCTT | AACCCAGAA  | CGCCAGCTGT | CCCCACGCTG | GACAGCACAC  | 26400 |
| TGAGGTTTGG | CGTGCAAGCC | GATCCGTGCA | CTTGCACTAC | TCCGGTTTTA | GTGGCACTCT  | 26460 |
| TAATGTCTTC | ATTGACCTTC | CTGATTTTAG | ACAGGAGGGT | CACGTCCACC | CTGACCCCAT  | 26520 |
| AGTGAAATC  | CACAGGCATG | ATTGCGGGCC | TAGACGCACA | GAGAAATCAC | AGGAAAGCTG  | 26580 |
| CGCGCACACT | GGGTGATCTG | GAGACGATAG | ACTGCCTTAA | ATAGAACTTT | TAGGGGAGGT  | 26640 |
| GGAAGTGTGC | GACATGGACA | GGTTAACCTT | CACAAATCGT | CAGTCACACA | CGTGGTGTAA  | 26700 |
| TCAGAAATTG | CTCGCTCAAA | AAAATTACAA | GCCTTGAAAC | TGCCGSTGTA | TGAGAGGGGG  | 26760 |
| CACGCTTCTG | GCGGAGGCGT | GCCAAATATG | GGAGGAACGA | AAATATCAGG | CAGAACTCTG  | 26820 |
| TCAGCGGTGG | CTTCCAGGAA | CTTCCGGATG | TCCACCACGT | TAACAAGCGT | CACCCCGGCC  | 26880 |
| GCCTTGGCCT | GGATAAACCG | AATCTCAATA | TTCACTGCCT | CCCTGAACAG | CGCCTGGACC  | 26940 |
| TCTGGSTGAC | TGGSTTTTTC | CTGTATCTCC | ACCATAGTGT | TGTACAACAT | ACTGGCGGCC  | 27000 |
| TTGGTGTGCA | GCAGCTCGTC | CCTGSAATG  | TATCGTTTGG | CAAGGCACAC | CCCGGGCATG  | 27060 |
| ATGCCTCGCA | CCCTGCACAA | ACTGATAGAG | TAGAAGGAGC | TAAATAAGTA | TATCCCTCTC  | 27120 |
| ACAATCAAAA | ACATCAGAAT | CTTCTGAGCT | TTGGTGGTGG | CCTTACGCAC | CCTGGAGTGA  | 27180 |
| AGCCACTCCA | GCTTCTCGCA | AAGGGCGGGG | TCCAAAATGA | TCTTGGCAGC | ATATGCTAGA  | 27240 |
| AGTTCCGCTC | GACTGTTGTT | GAAAAATATC | TTCAAGATAT | TGSCATACAC | GACACCGTGG  | 27300 |
| ATATTCTCCA | TGGCAACCTG | TTGGGCATAA | TAGTGGGCCA | CGTCGTGGCT | GTTAAAATTT  | 27360 |
| GTGACAAGGT | CCTCAATGTT | AAAGTTAACT | AGGCSTTCGG | CCATTCCCAA | AAACGTAAAC  | 27420 |
| AAAAATCTAT | AAAAGTCCTT | GTCGGCATCG | CTGACCTGGT | GCACGTGGGA | AACATCAAGG  | 27480 |
| TGCAGGGGTA | TCTGGCTAGG | AAACCATCGG | TTCTGCCAAG | TCTCGCGCGT | TAGCGCCAAA  | 27540 |
| AATCCGTCTG | GATCGCTTGT | ATACAGAAAT | CGATCAACTG | AATCCATTGG | CCTCACCCGG  | 27600 |

|            |            |             |             |             |            |       |
|------------|------------|-------------|-------------|-------------|------------|-------|
| CTTGCAGAGA | CCTACCTACT | GACAGACCAG  | GCACTCGGGG  | TCTGCCGCGC  | AGGACTCCTC | 27680 |
| CTCCGGGTTT | TTAGGTCCGG | GTAACCACGC  | CCCATCTTGT  | TTCAATCCAG  | AGTGAGGCGG | 27720 |
| TGACCCTGGA | TCTGCCAGGC | ACTGAAGAGC  | CGTCAGACTA  | GATTGCTTCT  | GAACCTTACA | 27760 |
| GTAGTACATG | AGGCTTTTAA | GACCAAGCCT  | GATCCCATGT  | AGCAGCAGGT  | CCCTAAGATA | 27800 |
| GCTCGCATTC | CTGACTCTGT | CCTCCTTGAG  | GAAGAAGCTC  | ATGGACTGGC  | TCTGGTCTAC | 27840 |
| AAACGGCGCC | CTGGCAGGAG | CCCTGTCCAG  | TAGCTTAAAT  | GGACAGTAAT  | CAAGGCTGT  | 27880 |
| TAGGAATACC | CTATATCTTT | CCCTGTGATG  | CTTGGGGAAC  | GTGGAAACGT  | CCCCACCATA | 27920 |
| CTGTCTAACC | ACCCGAAGGT | CGTCGGGGAG  | AACCTTCTTA  | AAAAAAGTCA  | CATTGGGGCT | 27960 |
| CAACACCTCT | TCTTTATTGG | TGACCTTGGA  | AGATATATTA  | GCAAAAAAGG  | GGTACACAGA | 28000 |
| CTCGGCATAG | CCAGTTACTT | GCGAGGTCCC  | AGCCGTCGGC  | ATCACCGCCA  | GAAACTGAGA | 28040 |
| ATTGAATATG | CCATGCTCGG | CAATGCTCTT  | TCCCAACGCG  | TCCAGCGAT   | GGCGTGCTAC | 28080 |
| AAACGAAGCA | TCCTCCCCCT | CCCATGTTTG  | CCAATGAAAC  | CTGCCCTTGG  | CGAAGTTACT | 28120 |
| GACCTCCAG  | CCATGAAATG | GGACACCCTG  | TCCCTCCAAA  | ACAAGGTTGT  | GACTAGTCTC | 28160 |
| CACCGCGGTG | TAGTACATAG | ACTGGAATAT  | ATTCTTGTCT  | AACTCAGCGC  | TCTCAGCATC | 28200 |
| GAGGTACCCG | TACCCCAATT | CCGCAAACAC  | ATCCGCCAAC  | CCCTGAACAC  | CAATCCCCAT | 28240 |
| AGACCTCTCC | TTTTGACCTC | GCTCGACCCC  | CGGTGTTGGA  | TGGGAACCAC  | CCAGAATGCA | 28280 |
| GGCGTTGATG | ACGAGGACTG | CCACCCTTAC  | TGCGTCGCCC  | AAGGCTCCAA  | AACAAAAAAA | 28320 |
| CGGCCTGTTG | GCGTCCGTGG | TGCCAACCCCT | CGCGCTTTCA  | ACAGTTCTCA  | GACACTTTGG | 28360 |
| AAGGCAGATA | TTTGCCAGGT | TGCACACCGA  | AGTGTCTCTT  | CCTGSCAGTT  | GGACTATCTC | 28400 |
| TGCACACAAG | TTTGAGCAGT | TAATGGCCAT  | GCCCTGAGTG  | TGGTCCAGT   | GGTGTTCATT | 28440 |
| GAGCGCTTCT | TTTAAAGCA  | CGTACGGTGA  | GCTGTCTCTT  | ATGATGGTGT  | GGATAAGAGT | 28480 |
| GAACATCATA | GACTTCAACG | GCATGCAACT  | AACGTACTTT  | CCAGCCCCGA  | CCAGGCGCTC | 28520 |
| GTATTGCTTA | TGAAACGCAG | CACCGTATAG  | CTTAATCAAA  | TTGGGGGCGG  | TGGCTGGATC | 28560 |
| GAACAAATAC | CATAACTTGG | ATGGGTCTCT  | TTCAATACATC | CTGAAAAACA  | ATGTTGGGAT | 28600 |
| GCACACGCCC | TGAAAGAGAC | TGTGACATCT  | GTCGGGATTC  | TCCGGTAGTT  | TGGGCTTCAA | 28640 |
| AAAATCACAG | ATTTGACTGT | GCCAGAGTTC  | CATGTATGCG  | CTCGCCCAAA  | CGGGCTTGAT | 28680 |
| GTTATTGTCA | TTGAAATAAT | GAACCTGGGC  | ATCCACCAGT  | TTGAGGCAAC  | TGGCTATGTT | 28720 |
| CTTTTGGTGG | GAGAAATGAC | TAACATCCAG  | ACCCACGCTT  | GACTTACTGG  | CCAGCAACGG | 28760 |
| ACTCATATCG | TGCTACAGGG | CGTCCAAAGT  | ACCCGACTCA  | TTCAATCATGG | AGGGCTGCAG | 28800 |
| AATAAAACAG | CTGGCGAGTT | GTCGCGCTTC  | GACTCCAGCT  | GAGCGCAGTA  | TTGGCGTGSC | 28840 |
| GCAGCACACG | TGCTGCGCAG | CGAGGTAGCC  | AAAAACGTAC  | TCCACTATAG  | CCATCTCAGA | 28880 |
| TACAGACTTA | GCGTCCCTCA | TAAGGTCCCG  | CGCCAAACCA  | TACAGGCATT  | CATGCTCTAA | 28920 |
| GCACTGACAG | GCAACAAACA | CGGAAACCTT  | CATAAACATT  | TGCGCCACGC  | TTTCATAGAC | 28960 |
| AGGCTCTGTC | CCCATGGTCC | TTAGGACGTA  | AGTATCATAC  | AACCTCACGG  | CCCATAGGTA | 29000 |

|             |             |            |             |             |             |       |
|-------------|-------------|------------|-------------|-------------|-------------|-------|
| GCCACAGTTA  | AGTGTGTCTT  | CGTAAGCTTT | GGACCGTCTG  | TAGGCGCACA  | ACATATCTTC  | 29700 |
| CAAGGCATCA  | ATGTTCTTTT  | GAATAAACGA | TTCCACCCGA  | TGTCCCAACA  | CGCCTCGAAA  | 29760 |
| AATCCCAAGA  | TACTGCTTGA  | GAGTCGCTGG | GCACCTAGCC  | TCCATAATTT  | GGTGCCACAG  | 29820 |
| CCGCCCCGGC  | ATGGCATTGG  | CCCGCACCTC | CCACCCGACC  | CTAACCTTTA  | GAAAGTCTAT  | 29880 |
| GAGAGATTGG  | GCACACATAT  | CAAAATCCGA | CAATTGTCCC  | GCAGACACCT  | GAGACCCCGC  | 29940 |
| TGGCTCTGGT  | GGGACAGCTC  | CCAAGTGAAC | CTGACAAAAT  | GTCCGGACAG  | ACATGACCTT  | 30000 |
| ACAGAAACAC  | AGTCCAGGGG  | CCACACGCGG | CCTCAAAGTT  | CGCAAACACC  | AGTACAGGCA  | 30060 |
| AGGACGTGCC  | CTTCACGTTT  | AGACTTTGGT | GCACCGGATG  | AGAATCAAAG  | GGAACTGTGC  | 30120 |
| CCAGCGTACA  | AACCGCCCCA  | AAAACAAGCC | GATTTATATA  | CAGCTCGTGC  | CTCAGCTGAA  | 30180 |
| TATACTTGGT  | CCGGATTACA  | TCCGTAAAGT | GATCCTTTAT  | CATGGCCACA  | ACCTCCGCAA  | 30240 |
| AGCCCTTCCC  | AGACTGGAAA  | AACGTCAGCG | CCATAGATGG  | TCTCTGGTTC  | ACACGGAGAT  | 30300 |
| AAACCAACGA  | GGCATAAATA  | GTAACGTTTA | GGCCTGCCGG  | TTCCCGGCGC  | TGGACCATGG  | 30360 |
| GACATGACTC  | ATCCAAATCA  | ACTAGCATAT | CACAAGGGAG  | GGTCAAGCCT  | ACGTGTGCAAC | 30420 |
| GGGGCTCCTC  | CCGGGCCAAC  | CCAACTCCCT | TCATGGCGGA  | GGTGACCTTG  | GTCAAGGAAGG | 30480 |
| TACTGTGGAC  | ACTCTGGACC  | ATTGGACCTA | CTGGGGTAAG  | GAGGGTATGA  | AACTCCCTAG  | 30540 |
| TGTCCATGAG  | TTCACTCAAG  | TTAGGGATGA | AATCCGCCAG  | GCCGGATCCA  | CTTCCGTACC  | 30600 |
| ACACACCGGC  | CACCTTTGTGA | GTCTGTGGCG | CTTTTGCCGC  | TTCCATTCCA  | GAGAGCATAA  | 30660 |
| ACAGGGACCT  | GGGTGTTAGC  | AGCATATCCA | TAGACGAGCC  | GTTGTCTCTC  | TGCTTGAATG  | 30720 |
| AAATAAAAA   | GGTTCCGAGA  | GGCTCCTGGG | GACTAAAGGT  | CTGTGAATAC  | ACGAGGAAAT  | 30780 |
| CTCCATAGGT  | CGGCTGCCTA  | AACGGCGCCT | GCCGCAAGGC  | CTCATGCAGC  | GAGCCAAACG  | 30840 |
| TGGGTCTGT   | GGAGGCGGCA  | TATTTAGAGA | GTAAATCCCG  | CACCCCCCTG  | GCAAACTCCG  | 30900 |
| GTCTCTAGT   | GAGGGATACC  | CGGTGAGTTG | GTGGAGGTAA  | AAGACCCAAC  | ACTTGCCTAC  | 30960 |
| CCAGGCGAGC  | CGCATTTTCA  | GCCTGCACCT | TCATATCCAC  | GCCGGCAATG  | GACGGCACAG  | 31020 |
| ACGCTCTTGA  | AAAGCTTACC  | AAAGGCCTGA | GTGGGGGAGG  | CGGGAGCCTT  | CACCAGACAA  | 31080 |
| AGCTGTTGAT  | GGAATTTCAA  | CTCCGAGGAC | TGCCGGTGCC  | TGCCCTCTTA  | AACAGCAGCA  | 31140 |
| CAACAGAGCA  | GTTTTTAAAT  | ACTGTTGCCC | AACTGCCGAC  | GGACCTATCA  | AAATTTATAC  | 31200 |
| GCGACTATCG  | CCTGTTCCGA  | CTGTTTCGGG | CGGCGTATTT  | TTTAGAACCC  | CCTTCTAGCA  | 31260 |
| TGACCCCCCT  | TGAGGCAGCG  | CGCGCTCTTG | GACGCCTGGT  | TGATATATTA  | TCATCACAAC  | 31320 |
| CACCGCAGAA  | CACCGCACCG  | GCGCAGCCAC | CCACCTCCGA  | CGACACCCCTG | AATAACTGTA  | 31380 |
| CATTGCTCAA  | ACTACTAGCC  | CACACGCGGG | ATCAGATAGC  | AGTTTTCAAA  | ACCCCCGCTC  | 31440 |
| TCCCTCCCGT  | GCCACCTGGA  | ATCATCGGCC | TGTTCCACATG | CGTGGAAACAG | ATGTACCAAG  | 31500 |
| CATGTTTTCA  | GAAATACTGG  | GCAGCTGCAC | TACCCCCAAT  | GTGGATACTG  | ACATACGACC  | 31560 |
| CTCCCACTTC  | TCCGTTACAG  | GACTGGCTTA | TAGTCGCCCTA | TGTTAACCAAG | GAAGGACTGC  | 31620 |
| TACTCCCCCTC | TGGCATACCC  | TGGGAGGAGG | TGTTAGCCAA  | AACATTAGTA  | ACAGAAACACC | 31680 |

|             |            |            |            |            |             |       |
|-------------|------------|------------|------------|------------|-------------|-------|
| ACGAGTTGTT  | CGTATCGCGG | TCGAATTGGA | CCGAGACCGG | CGTCACCATG | CCCGTATCCA  | 31740 |
| AAGAACGGCG  | CCTCGCCATC | TACCGGGTGT | TCGCCAAGGG | TGAGGTGGTG | GCGGAAAATA  | 31800 |
| CTCCCATTTT  | TGCCTTCACC | GACGTGGAA  | TATCCACACT | CAAAACCCAC | TATCTGTTCA  | 31860 |
| TCTATGATTT  | TATCATAGAG | GCATTATGCA | AGAGCTACAC | ATACTCATGC | ACCCAGGCCC  | 31920 |
| GCCTGGAATC  | CTTTTGTAGC | CGAGGTATAG | ACTTCATGAC | TGACCTAGGT | CAGTACCTAG  | 31980 |
| ATACCGCTAC  | TAGCGGCAAG | CAGCAGCTGA | CGCACAGCCA | AATAAAGGAA | ATCAAATACA  | 32040 |
| GGCTGCTAAG  | CTGCGGTCTC | TCGGCTTCCG | CGTGTGATGT | TTTCAGAACT | GTGATCATGA  | 32100 |
| CCCTCCCATC  | TGACCGGACC | CCCAACCTCG | CTAACCTGTC | CACGTTTATG | GGGATGGTTC  | 32160 |
| ACCAACTGAC  | CATGTTCCGA | CACTATTTCT | ACCGGTGCCT | GGGCAGCTAC | AGTCCACCGG  | 32220 |
| GCTTGGCCTT  | CACAGAATTG | CAAAAGATAC | TGACACGCGC | CAGCGCGGAG | CAAACGGAAC  | 32280 |
| GTAACCCGTG  | GAGACATCCG | GGTATCTCGG | ACATTCCACT | GCGTTGGAAA | ATATCGCGTG  | 32340 |
| CTCTAGCATT  | CTTCGTCCCT | CCGGCCCCCA | TAAACACTTT | GCAGCGCGTG | TACGCCGCGC  | 32400 |
| TGCCCTCGCA  | ACTCATGCGG | GCCATCTTCG | AGATCTCGGT | CAAGACCACA | TGGG3AGGCG  | 32460 |
| CCGTACCGGC  | AAACCTGGCG | CGCGACATTG | ACACAGGACC | GAACACACAA | CATATCTCCT  | 32520 |
| CCACATCACC  | GCCCACCTTC | AAGGATGTTG | AGACATACTG | TCAAGGTCTG | CGG3TGGGAG  | 32580 |
| ACACGGAGTA  | CGATGAGGAC | ATTGTGAGAA | GCCCCGTCTT | TGCAGACGCG | TTTACCAAGA  | 32640 |
| GTCACCTGTT  | GCCTATACTG | CGCGAGGTTG | TGGAAAACCG | CCTGCAGAAA | AACAGAGCTC  | 32700 |
| TGTTTCAGAT  | AAGATGGCTG | ATAATATTTG | CTGCCGAGGC | GGCAACCGGG | CTCATCCCTG  | 32760 |
| CCAGGCGCCC  | GCTAGCCAGA | GCCTACTTCC | ACATCATGGA | CATTCTGGAG | GAGAGACATT  | 32820 |
| CCCAAGACCG  | CCTATACAAC | CTTTTGGACT | GATCCAGGA  | GCTCTTCACC | CACATCAGGC  | 32880 |
| AGGCTGTTCC  | AGACGCACAG | TGTCCGCACG | CCTTTCTACA | GTCCTGTTC  | GTCTTTCAAT  | 32940 |
| TCCGCTCTTT  | CGTACTCAAA | CACCAGCAGG | GTGTAACCTT | GTTTCTAGAT | GGCTTGCAGA  | 33000 |
| CATCCCTCCC  | CCCG3TGATA | AGTGTGGCCA | ACCTTGGAGA | CAAGCTGTGT | CGTCTCGAGT  | 33060 |
| TCGASTACGA  | CAGCGAGGGC | GACTTCGTGC | GCSTGCCAGT | TGCACCGCCA | GAACAACCAAC | 33120 |
| CGCACGTACA  | TCTGTCCGAT | TTCAAGAAGA | CAATACAGAC | CATCGAACAG | GCCACCAGGG  | 33180 |
| AGGCCACCTT  | AGCCATGACA | ACAATCGCAA | AGCCAATATA | CCCCGCTTAC | ATCCGGTTAC  | 33240 |
| TGCAGCGGCT  | AGAATATCTT | AACAGACTCA | ACCACCACAT | TCTCAGGATT | CCCTTCCCAC  | 33300 |
| AGGACGCCCT  | TTCTGAACTC | CAGGAAACCT | ACCTGGCGGC | GTTTGCACGG | TTGACAAAAT  | 33360 |
| TGGCAGCGGA  | CGCAGCAAAC | ACTTGTAGCT | ACTCCCTCAC | CAAGTACTTT | GGAGTTTTAT  | 33420 |
| TCCAAACACCA | GCTGGTCCCC | ACGGCCATCG | TTAAAAAAGT | GCTACATTTT | GACGAGGCTA  | 33480 |
| AAGATACCAC  | AGAAGCCTTT | TTACAGAGCC | TGGCACAACC | CSTAGTGCAG | GGACAACGGC  | 33540 |
| AGGGGGCGGT  | TGGCGGGTCC | GGTGTCTCTG | CGCAGAAAGA | ACTTGAGCTC | TTGAACAAAA  | 33600 |
| TAAACCCACA  | GTTTACAGAC | GCTCAGGCTA | ACATTCTCTC | ATCTATTAAA | CGTTTCATATT | 33660 |
| CAATAAATA   | TGACSTCCCT | GAGGTCTCAG | TCGACTGGGA | AACGTACTCC | CGGTCTGCCT  | 33720 |

|             |            |            |             |            |            |       |
|-------------|------------|------------|-------------|------------|------------|-------|
| TCGAGGCACC  | GGACGACGAA | CTCCGTTTTG | TCCCACTGAC  | GCTGGCAGGC | CTCCGGAAAC | 33780 |
| TGTTTGTGGA  | ATAGAGGCCA | TGGCAGCCCA | GCCTCTGTAC  | ATGGAGGGAA | TGGCCTCCAC | 33840 |
| CCACCAAGCT  | AACTGTATAT | TCGGAGAACA | TGCTGGATCC  | CAGTGCCTCA | GCAACTGCGT | 33900 |
| CATGTACCTG  | GCGTCCAGCT | ATTATAACAG | CGAAACCCCC  | CTCGTCGACA | GAGCCAGCCT | 33960 |
| GGACGATGTA  | CTTGAACAGG | GCATGAGGCT | GGACCTCCTC  | CTACGAAAT  | CTGGCATGCT | 34020 |
| GGGATTTAGA  | CAATATGCCC | AACTTCATCA | CATCCCCGGA  | TTCTCTCCGA | CAGACGACTG | 34080 |
| GGCCACCAAG  | ATCTTCCAGT | CTCCAGAGTT | TTATGGGCTC  | ATCGGACAGG | ACGCGGCCAT | 34140 |
| CCGCGAGCCA  | TTCATCGAGT | CCTTGAGGTC | GGTTTTGAGT  | CGAAACTACG | CGGGCAGGCT | 34200 |
| ACAGTACCTG  | ATCATTATCT | GCCAGTCCAA | AGCCGGAGCA  | ATCGTCGTCA | AGGACAAAAC | 34260 |
| GTATTACATG  | TTTGACCCCC | ACTGCATACC | AAACATCCCC  | AACAGTCCTG | CACACCTCAT | 34320 |
| AAAGACTAAC  | GACGTTGGCG | TTTTATTACC | GTACATAGCC  | ACACATGACA | CTGAATACAC | 34380 |
| CGGGTGCTTC  | CTTTACTTTA | TCCCACATGA | CTACATCAGC  | CCAGAGCACT | ACATCGCAAA | 34440 |
| CCACTACCGC  | ACCATTGTGT | TCGAAGAACT | CCACGGGCCC  | AGAATGGATA | TCTCCCGCGG | 34500 |
| GGTGGAAATCA | TGCTCCATCA | CCGAAATCAC | GTCCCTTTCT  | GTATCCCCCG | CGCCTAGTGA | 34560 |
| GGCACCATTG  | CGCAGGGACT | CCACCCAATC | ACAAGACGAA  | ACGCGCCCCG | GCAGACCTCG | 34620 |
| CCTCGTCATT  | CCTCCTTACG | ATCCGACAGA | CCGCCCCACGA | CCGCTTCACC | AAGACCGCCC | 34680 |
| GCCAGAGCAG  | GCAGCGGGAT | ACGGTGGAAA | CAAAGGACGC  | GGCGGTAAAC | AAGGACGCGG | 34740 |
| CGGAAAGACG  | GGACGTGGCG | GAAATGAAGG | ACGCGGTGGC  | CACAGCCAC  | CAGACGAGCA | 34800 |
| CCAGCCCCCA  | CACATCACCG | CGGAACACAT | GGACCACTCC  | GACCGACAAG | GCGCCGATGG | 34860 |
| AGACATGGAT  | AGTACACCCG | CAAATGGTGA | GACATCCGTT  | ACGGAAACCC | CGGGCCCCGA | 34920 |
| ACCCAATCCC  | CCAGCACGGC | CTGACAGAGA | GCCACCGCCC  | ACTCCCCCGG | CGACCCCAGG | 34980 |
| CGCCACAGCG  | CTGCTCTCTG | ACCTAACTGC | CACAAGAGGG  | CAGAAACGCA | AATTTTCTCT | 35040 |
| GCTTAAAGAA  | TCTTATCCCA | TCGACAGCCC | ACCTCTGAC   | GACGATGATG | TCTCCACGCT | 35100 |

## (2) INFORMATION FOR SEQ ID NO:20:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 32207 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| CTCCCAACAA | ACGGCTCCGG | ATACTGAAGA | TATTTGGATT | GACGACCCAC | TCACACCTTT | 60  |
| GTACCCACTA | ACGGATACAC | CATCTTTTGA | CATAACGGCG | GACGTCACAC | CTGACACAC  | 120 |
| CCACCCCGAG | AAAGCAGCGG | ACGGGGAGTT | TACCAACAAG | ACCACAAGCA | CGGATGCGGA | 180 |



|            |             |             |            |             |             |      |
|------------|-------------|-------------|------------|-------------|-------------|------|
| CAGGTATGCC | AGCGCCAGTC  | AGGAATCGCT  | GGGCACCCTG | GTCTCGCCAT  | ACGATTTTAC  | 240  |
| AAACTTGGAT | AACTGCTGG   | CAGAGCTGGG  | CCGGTTGGGA | ACGGCACAGC  | CTATCCCTGT  | 300  |
| AATCGTGGAC | AGACTAACAT  | CGCGACCTTT  | TCGAGAAGCC | AGCGCTCTAC  | AGGCTATGGA  | 360  |
| TAGGATACTA | ACACACGTGG  | TCCTAGAATA  | CGGTCTGGTT | TCGGGTTACA  | GCACAGCTGC  | 420  |
| CCCATCCAAA | TGCACCCACG  | TCCTCCAGTT  | TTTCATTTTG | TGGGGCGAAA  | AACTCGGCAT  | 480  |
| ACCAACGGAG | GACGCAAAGA  | CGCTCCTGGA  | AAGCGCACTG | GAGATCCCGG  | CAATGTGCGA  | 540  |
| GATCGTCCAA | CAGGGCCGGT  | TGAAGGAGCC  | CACGTTCTCC | CGCCACATTA  | TAAGCAAGCT  | 600  |
| AAACCCCTGC | TTGGAATCCC  | TACACGCCAC  | TAGTCGTGAG | GACTTCAAGT  | CCCTGATACA  | 660  |
| GGCATTCAAC | GCCGAAGGGA  | TTAGGATCGC  | CTCGCGTGAG | AGGGAGACGT  | CCATGGCCGA  | 720  |
| ACTGATAGAA | ACGATAACCG  | CCCGCCTTAA  | ACCAAATTTT | AACATTGTCT  | GTGCCCCCCA  | 780  |
| GGACGCACAA | ACCATTCAAG  | ACGGCGTCGG  | TCTCCTCAGG | GCCGAGGTTA  | ACAAGAGAAA  | 840  |
| CGCACAGATA | GCCCAGGAGG  | CTGCGTATTT  | TGAGAATATA | ATCACGGCCC  | TCTCCACATT  | 900  |
| CCAAACCCTT | CCCCAATCGC  | AACAGACGTT  | CGAAGTGCTG | CCGGACCTCA  | AACTGCGCAC  | 960  |
| GCTCGTGGAG | CACCTGACCC  | TGGTTGAGGC  | GCAGGTGACA | ACGCAAACGG  | TGGAAAGTCT  | 1020 |
| ACAGGCATAC | CTACAGAGCG  | CTGCCACTGC  | TGAGCATCAC | CTTACCAACG  | TGCCCCAACGT | 1080 |
| CCACAGTATA | CTGTCTAACA  | TATCCAACAC  | TCTAAAAGTT | ATAGATTATG  | TAATTCCAAA  | 1140 |
| ATTTATAATA | AACACCGATA  | CACCTGGCCCC | ATATAAACAG | CAGTTTTTCAT | ATCTGGGGGG  | 1200 |
| TGAAC TGGA | TCTATGTTCT  | CCCTTGACTG  | GCCTCACGCA | CCTGCAGAGG  | CGGTAGAGCC  | 1260 |
| ACTACCCGTG | CTGACTTCTC  | TGCGAGGTAA  | AATCGCAGAG | GCGCTGACGC  | GTCAAGAAAA  | 1320 |
| CAAAAACCTT | GTAGATCAAA  | TTCTAACCGA  | CGCCGAAGGC | CTCCTTAAGA  | ACATTACCGA  | 1380 |
| TCCAAACCTC | GCACACTTCC  | ACGCCCAGGC  | CGTATCAATT | CCAGTSTTAG  | AAAAC TACGT | 1440 |
| ACATAACCTG | GGGCTCCTTC  | TCAAGGGCGA  | AAAGAGCGAG | AGGTTCTCCC  | GGCTGAAGAC  | 1500 |
| CGCCATCCAA | AACTTGTTAT  | CCTCCGAATC  | ATTTATCACC | GTGACCTTAC  | ACAGTACAAA  | 1560 |
| CCTTGGAATC | CTAGTTACCA  | ACGTACCAAA  | ACTTGSTGAG | GCGTTACCGG  | GGGGCCCCGA  | 1620 |
| CCTCCTGACA | AGCCCCGTCC  | TGAGACAGTC  | CCTTTCCACC | CTGTGCACAA  | CCCTGCTGCG  | 1680 |
| AGATGCCCTG | GACGCCCTGG  | AAAAAAAGGA  | TCCGGCCCTT | CTTGGTGAGG  | GGACCACTTT  | 1740 |
| GGCGCTGGAG | ACACTCCTAG  | GATACGGGTC  | GGTGCAGGAC | TACAAGGAGA  | CGGTACAGAT  | 1800 |
| AATATCCAGC | CTTGTGGGCA  | TCCAAAAATT  | AGTCAGGGAC | CAGGGCGCGG  | ACAASTGGGC  | 1860 |
| CACCTCCCTG | ACAAAGGCTAA | CTGACCTCAA  | ATCAACTCTG | GCCACGACCG  | CCATCGAGAC  | 1920 |
| GGCTACGAAA | CGGAAACTAT  | ACAGATTGAT  | CCAAAGGGAC | CTCAAGAGG   | CTCAAAAAAC  | 1980 |
| CGAGACCAAT | CGGGCCATGG  | AGGAATGGAA  | GCAGAAAGTA | CTGGCTCTTG  | ACAATGCGTC  | 2040 |
| TCCGGAACGT | GTGCCCACCC  | TCTTGCAACA  | GGCTCCCATC | GCGAAGGCTA  | GAGAGTTTGC  | 2100 |
| AGAGAAGCAC | TTCAAAATAC  | TACTCCCGGT  | ACCCGCGGAC | GCCCCCGTCC  | AAGCTCTCTC  | 2160 |
| AAAGCCGATG | GAATACAGCG  | CCAGCCCCCT  | CCCGGACCCA | AAGGATATAG  | ACAGAGCTAC  | 2220 |

|             |            |            |            |            |             |      |
|-------------|------------|------------|------------|------------|-------------|------|
| ATCCATCCAC  | GGGGAACAGG | CSTGGAAGAA | GATACAGCAG | GCGTTCAAGG | ATTTCAACTT  | 2280 |
| CGCCGTCCCTG | CGGCCCGCTG | ACTGGGATGC | CCTGGCAGCG | GAGTACCAAC | GCCGTGCTTC  | 2340 |
| GCCCCCTCCG  | GCGGCCGTGG | GTCCAGCGCT | CTCAGGGTTC | CTGGAGACGA | TCCTAGGGAC  | 2400 |
| GCTGAACGAC  | ATCTACATGG | ATAAGCTCCG | CTCCTTTCTG | CCCCACGCGC | AGCCTTTTCA  | 2460 |
| GGCGCCGCCC  | TTGACTGGC  | TAACGCCGTA | TCAGGACCAA | GTCAGCTTTT | TCTTGCGCAC  | 2520 |
| CATAGGGCTG  | CCGCTGGTGC | GAGCGCTGGC | CGACAAGATC | AGCGTGACAG | CACTGAGGCT  | 2580 |
| TAGCCACGCG  | CTCCAGTCCG | GCGATTTGCA | GCAGGCCACG | GTGGGCACGC | CCCTGGAGCT  | 2640 |
| CCCTGCCACA  | GAGTACGCGC | GCATCGCCTC | CAACATGAAG | TCCGTGTTCA | ACGACCCACG  | 2700 |
| ACTTCAGGTG  | CGATCAGAGG | TCGCGGATTA | TGTGGAGGCC | CAACGAGCCG | ACGCACACAT  | 2760 |
| GCCACACGTC  | CCACGTCCAA | AGATACAGGC | ACCAAAGACT | CTGATTCCAC | ATCCGGACGC  | 2820 |
| AATCGTCGCG  | GACGGACTAC | CCGCCTTTCT | TAAGACGTCC | CTACTGCAGC | AAGAGGCCAA  | 2880 |
| ACTTCTGGCG  | CTACAGCGGG | CGGACTTCGA | GTCGCTCGAG | AGCGACATGC | GCGCCGCAGA  | 2940 |
| GGCCGAGAGA  | AAAGCATCGC | GCGAGGAAAC | CCAGCGCAAA | ATGGCACACG | CCATCACTCA  | 3000 |
| GCTCTTACAG  | CAGGCACCCA | GTGCGATCTC | GGGGCGCCCG | CTATCCTTAC | AGGACCCGGT  | 3060 |
| GGGCTTCTCT  | GAGGGCATCA | TATACGACAA | GGTCTTGAG  | CGCGAATCCT | ACGAGACGGG  | 3120 |
| TCTCGAGGGA  | CTGTCTCTGG | TCGAGCAGAC | CATCAAGTCC | ATCACCGTAT | ACGCTCCCGT  | 3180 |
| AGAGGAGAAG  | CAAAGAATGC | ACGTGCTGCT | GGACGAGGTG | AAAAAGCAGC | GAGCAACAC   | 3240 |
| TGAGACCGCT  | CTCGAGCTAG | AGGCCGCGGC | TACGCACGGC | GACGACGCTA | GACTCCTGCA  | 3300 |
| GCGAGCGGTC  | GATGAGCTGT | CACCGTTGCG | CCTTAAGGGG | GGGAAGGCGG | CGGTGGAAATC | 3360 |
| CTGGCGGCGAG | AAAATCCAAA | CCCTGAAATC | CCTGGTACAG | GAAGCGGAGC | AGGCCGGCCT  | 3420 |
| CCTGTTGGCC  | ACCATAGACA | CGGTGGCCGG | CCAGGCCACG | GAGACCATAT | CACCATCCAC  | 3480 |
| ACTCCAGGGA  | CTGTACCAAC | AGGGACAGGA | GSCCATGGCG | GCCATTAAGC | GGTTTAGGGA  | 3540 |
| CTCGCCCCAG  | CTAGCTGGCC | TGCAGGAAAA | GCTGSCCGAG | CTACAGCAGT | ACGTCAAGTA  | 3600 |
| CAAGAAAGCAG | TATCTGGAAC | ACTTTGAGGC | CACCCAAAGC | GTAAGTGTTA | CAGCCTTTCC  | 3660 |
| GCTCACACAG  | GAGGTTACGA | TCCCAGCCCT | GCATTACGCG | GGACCTTTCC | ACAACTTGGA  | 3720 |
| GCGGCTCTCA  | CGATACCTAC | ACATCGGCCA | GACGCAGCCG | GCTCCGGGAC | AGTGCTCTCT  | 3780 |
| GACACTTCCC  | ACATTCGACC | CCACGCGCCC | GGCCTGCGTC | CCAGCCGGCG | GCCACGAACC  | 3840 |
| CCCGTTGCAC  | AGACAGGTGG | TGTTCTCCAG | CTTTTGGAG  | GCCCAGATCC | GATTAGCGTT  | 3900 |
| GTCCGTAGCG  | GGCCCCCTGC | CTGGACGGGG | TCTGCCCGGA | ACACTGCAGA | TCCGAAGGGG  | 3960 |
| CGTGAGGGCT  | GCCGCTTCTT | TCTCTCACCA | GTGGGACGAG | ATATCTCGCC | TCCTTCCAGA  | 4020 |
| GGTACTGGAC  | ACCTTTTTTC | ACAAGCGCGC | CCTTCCCGCA | GAGTCTTCTT | CCAAATGCTTT | 4080 |
| CCTGGCCATG  | TGCGTATTGA | CGCACCTTGT | CTACCTAGCT | GGGCGCGCCG | TCTTGCGCCC  | 4140 |
| ACGGGAGCCG  | GAGCACGCGC | CCCCGGACGC | GTACCCAAAG | GAGGTGGCGC | TGGCCCCGCG  | 4200 |
| CGACCTGACC  | TACCTTCTAC | TGGCCATGTG | GCCATCTTGG | ATCTCGGCCA | TTTTGAAACA  | 4260 |

|            |            |            |             |             |             |      |
|------------|------------|------------|-------------|-------------|-------------|------|
| GCCTTCGCAC | GCGGAGGCGG | CGCACGCATG | TCTTGTCAAG  | CTGCCAACAA  | TGCTCAAGGT  | 4320 |
| TGTGCCGTAC | CTCACGCTGG | AAGCCTCAGC | TGGACCACTG  | CCGGCGSACA  | TGCGGCACCTT | 4380 |
| CGCCACGCCA | GAAGCGCGTC | TGTTTTTCCC | CGCGCGATGG  | CACCACGTCA  | ACGTGCAGGA  | 4440 |
| GAAACTGTGG | CTGCGTAATG | ATTTTATGTC | GCTGTGTCAAC | CGTTCCCCGG  | GGCGCGCGCG  | 4500 |
| CATAGCCGTC | TTGGTGTGGG | CCGTCACTTG | CCTAGATCCT  | GAGGTAATAA  | GGCAGCTGTG  | 4560 |
| GTCCACCTTG | CGGCCCTTA  | CTGCGGATGA | ATCCGACACG  | GCTTCTGGAC  | TGCTGCGGGT  | 4620 |
| GCTAGTAGAA | ATGGAGTTTG | GTCCGCCGCC | CAAGACGCCG  | CGGCGGGAGG  | CGGTGGCGCC  | 4680 |
| CGGCGCAACA | CTGCCACCGT | ACCCCTACGG | CCTTGCCACC  | GGCGAGCGGC  | TGGTGGGCCA  | 4740 |
| GGCGCAGGAA | CGCTCTGGCG | GCGCTGGCAA | GATGCCGGTG  | TCCGGGTTTG  | AGATAGTTTT  | 4800 |
| AGGCGCACTG | CTGTTCCGCG | CCCCCTACG  | CATTTTCAGC  | ACCGCATCAA  | CCCACAGGAT  | 4860 |
| CTCAGATTTC | GAGGGCGGTT | TCCAGATACT | GACTCCTCTC  | CTGGACTGTT  | GCCCAGATCG  | 4920 |
| CGAGCCATTC | GCCTCCCTGG | CGCGCGCACC | ACGAAGGACG  | GTGCCACTGG  | GAGACCCGTG  | 4980 |
| CGCCAACATT | CACACCCCGG | AAGAGATACA | GATCTTTGCG  | CGTCAAGCCG  | CCTGGCTTCA  | 5040 |
| ATATACCTTC | GCAAATTACC | AGATCCCCAG | CACCGACAAC  | CCGATACCGA  | TGTTTGTGCT  | 5100 |
| AAAGCCTAAC | AATAACCTTG | AAAACAGCTA | CATCCCTCGC  | GATCGCAAAG  | CGGACCCGCT  | 5160 |
| ACGACCATTG | TATGTAGTCC | CTCTGAAGCC | GCAGGGTAGA  | TGGCCTGAAA  | TAATGACCAC  | 5220 |
| AGCAACAACC | CCCTGCCGCC | TACCGACATC | GCCAGAAGAG  | GCGGGATCAC  | AGTTGGCCAG  | 5280 |
| ACTCCTTCAG | AGCCAGGTGA | GCGCCACATG | GTCTGACATC  | TTCTCCAGGG  | TTCCCGAGCG  | 5340 |
| CCTCGCTCCC | AATGCGCCTC | AGAAGAGTTC | CCAGACAATG  | TCAGAAATCC  | ACGAGGTCCG  | 5400 |
| CGCCACGCCG | CCACTCACAA | TCACCCCAA  | TAAACCGACC  | GGAACCCCTC  | ACGTCTCCCC  | 5460 |
| GGAGGCTGAT | CCAATAACAG | AACGCAACCG | CGGACAGCAG  | CCGAAGATTG  | TCCGGGACAA  | 5520 |
| CATGCCATAG | CGTATTCTCC | CGTCGCTACC | GACCCCGAAA  | CCCAGAGAGC  | CTAGAATCAC  | 5580 |
| GCTACCCACG | GCACTGCCCG | TTATATCACC | CCCAGCACAT  | CGCCCGTCCG  | CTATACCGCA  | 5640 |
| TCTGCCAGCA | CCGCAGGTAA | CGGAGGCCAA | AGGGGTTCTC  | CAAAGCAAAAC | GTGGAACTTT  | 5700 |
| CGTGCTGCCG | CCCGCCGCGG | TCATTGACCC | ACGGAAGCCC  | GTCTGGGCAC  | CGATCACGGG  | 5760 |
| ATATGAGAGG | ACGGCGCTCC | AGCCCCCCCC | GACTGAGGGC  | GAAGGGCGGC  | GCCCTCCCGA  | 5820 |
| CACGCAACCT | GTCACTTTAA | CCTTTCTGCT | CCCACCTACC  | GCACCCACTC  | CGGCAACTGT  | 5880 |
| AGCCCTAGAA | ACCAAAACAA | CTCCCCCATC | CACGCCCCCA  | CACGCCATAG  | ACATTAGCCC  | 5940 |
| ACCACAGACA | CCTCCCATGT | CCACCTCACC | TCACGCGAGA  | GACACAAGCC  | CCCCCGCAGA  | 6000 |
| AAAGCGGGCT | GCACCCGTCA | TTGAGTAAT  | GGCGCCCAAG  | CAACCGTCCG  | GAGAGGCAAG  | 6060 |
| AGTCAAGCCA | GTGGAGATCG | AACAGGGCCT | TTCCACACGC  | AATGAAGCCC  | CTCCCTTGA   | 6120 |
| ACGCTCGAAT | CACGCCGTGC | CGGCGTTTAC | CCCAAGGCGC  | ACCGTAGCCC  | GCGAAATCAG  | 6180 |
| GATCCCCCCC | GAGATAAAGG | CGGCTTGGGA | CCTGCACCGG  | GACATTCTCT  | TGCCCCACAG  | 6240 |
| CTCCCCGGAG | TCATCCCCAC | CGACTTCCCC | CCAGCCTATC  | CGCGTGGATG  | ATAAATCGCC  | 6300 |

|             |             |            |             |            |            |      |
|-------------|-------------|------------|-------------|------------|------------|------|
| TOTTCGCCAAC | CTCGTAGAGA  | GATACGCGCG | GGGTTTCCTG  | GACACGCCCT | CTGTAGAGGT | 6360 |
| GATGTCCCTG  | GAAAATCAGG  | ACATCGCCGT | GGACCCCGGA  | CTGCTAACCC | GCCGGATTCC | 6420 |
| ATCCGTGGTG  | CCCATGCCCC  | ATCCAATTAT | GTGGTCACCC  | ATAGTACCCA | TCAGTTTACA | 6480 |
| AAACACAGAC  | ATAGACACTG  | CAAAGATAAC | ACTGATTAGT  | TTTATTAGAC | GCATCAACAA | 6540 |
| AAAAGTGGCC  | GCCCTATCGG  | CGTCCCTGGC | GGAGACGGTT  | GACAGAATAA | AGAAGTGGTA | 6600 |
| CTTGTGACTC  | CACGGTTGTC  | CAATCGTTGC | CTATTTCTTT  | TTGCCAGAGG | GGGGTTTCTT | 6660 |
| CGCGTGGGCC  | ACCGCGGGGG  | CGGCCGTTTC | CGTCGTGGAT  | GAGAGGGTTG | TGAGAATGTC | 6720 |
| TGACGCCGGC  | GACAATGAAT  | GGGGACCAGA | GGACAGGGTG  | GTTATACTGC | TTCCCGAGAC | 6780 |
| CCCCAGTGAG  | TCCTGGCCCC  | CGGGCGTGGT | GCCGGATGCA  | GGGCCTGGCC | TCGAAGGCAC | 6840 |
| GGTGAACGTC  | CCCGCGTCGT  | AAGCCGACGC | CGCGGAACCT  | CGGTACAGCG | GCTCGCGCGG | 6900 |
| TTTCTGATCC  | CTAAGGGTCT  | GCAGATGATC | CCGCCTTTGA  | ATTCCACCCA | TCCTCTCTAG | 6960 |
| ATAGGCCTCA  | TAATAATGAT  | GGGCAATTAA | GAACACGAGA  | TAGTGTCTCT | TTTGCAAGAG | 7020 |
| GTATTCGGCC  | TGCGACATAT  | TTCCCTGATC | CAGGGTATTC  | ATGCGAGCCA | CCAGGGGATG | 7080 |
| GTGAGCGTAG  | TCATGATCCA  | GTGCTCCTG  | GATCACGGGG  | TCTCTCACCT | TAAAGTTGGA | 7140 |
| CATCTTCCAC  | ACAGGCGGGC  | GAAATAGCCT | CAGGAGGAAC  | ACTTCCCGCA | ACAGAACTCC | 7200 |
| AGCAGCTGTG  | AGGTGAGCTG  | AAGCAGTCCG | CGCACGTCAC  | GGTGCTTTAA | TAGGGCAGCC | 7260 |
| TCGCACTCGG  | GGTCCCAAG   | GCAAGGCACT | ACAAAACCTGA | CAGTTTGATC | TAGGTCTCGA | 7320 |
| ATGGCAAGGG  | CCGCGTTGTT  | AGCTAGAACA | GCCCTGATTA  | CGACGCGTGC | TAGGTCTCCG | 7380 |
| CGTCCGGTAA  | TATCGCACAG  | GGGATACACC | CTCATATGTT  | CGCTGCCACA | GTAAGAACAG | 7440 |
| TAGATCCTCC  | CCGTGGTCGC  | ACAGATGGTG | AACTGCTTCT  | CTTTCTCTGT | CCTGCTGAAA | 7500 |
| AACACGTTGG  | TGGGAGGAAA  | ATTGACAGTA | TGAAACTTGC  | CCCTGCCAAA | GTTAAGACAG | 7560 |
| TGTCCACACT  | CCATGCACAC  | AACCGCCCCA | GCGCAACGGC  | CCCGCTTGGC | AAGGSCCGCG | 7620 |
| CGGGCCACGC  | GAGAACAGAT  | GACGGGTATG | GACACGCAGG  | GGGAGAGAAC | ATTGTATGCC | 7680 |
| AGAAGCCTCC  | TGCCAAGGTT  | CCGCACGAGA | CCAGGTCCCT  | CCTGCTCCCA | GGGGGGCAGC | 7740 |
| ACTACGTGGC  | GGGACTTAAT  | AAGGCTCAAA | AAACACAGTG  | ACCCAAGCAT | GGCGTCGAAC | 7800 |
| GGGTTACCGC  | AGGGAACCGT  | AGGGGCGACG | CGCTCCAAAG  | CCTCCCGGAG | GCCGGTATCT | 7860 |
| GGCGCCCTTA  | TCCCGAGCCC  | GTTACCGTCT | TGGTCCGAG   | CCACACCGCG | ACGGGTGTGC | 7920 |
| GAGGGCACTT  | CCAGGAGGGG  | ACGACGCGGC | AACGGCCCCAT | GCCACTTCTT | CCTTAGCCAG | 7980 |
| GGTAGCGACG  | GTGGGGGGCTT | CGAACAGTAG | GTCACCTAACG | GAAAGCGAGA | GCAAAGCGCC | 8040 |
| AACAGCTTGC  | AGAGTTGGGC  | ACAGGCCTTG | GAAATGGAA   | GGACAGGTA  | TTTTGCCCAT | 8100 |
| ACGTGSCCGG  | GATCGCCCTT  | AGCATGGTGG | GCGGCTGGG   | CACGGGACAG | CGTCACACAA | 8160 |
| ACCCATACGT  | GGGCGCCAA   | CAGCTGTGTC | GCGGCACAAA  | TCTGCGCCTG | TTTGGCGACG | 8220 |
| GTCTCTGAGC  | CAGCGCGCAA  | CACGGCGATC | GCCTGCGCCA  | GCGACGGGCG | GTCCACACAG | 8280 |
| TGCCTGGCCC  | AGGAGGGCAT  | GTTTCCTGCG | AAACCCCGCT  | CCCGCAATAT | GACAAAAGCT | 8340 |

|            |             |            |             |             |             |       |
|------------|-------------|------------|-------------|-------------|-------------|-------|
| ACATATTCTT | CCACTGGCAC  | GCCATTCTCG | CCCTCGAACA  | CGCGGTGGGC  | CGTCAGCTGG  | 8400  |
| GCCTCATCCA | AACCAACCA   | AGACACAAGA | AAGCGATCCC  | AGCGCTGATC  | CAGGGCCATG  | 8460  |
| ACCTTCTEAC | CAGCGCGACC  | GCACGGCCTA | AGCTCCACTG  | AAAGGCGCCC  | AGAATCCGCA  | 8520  |
| CGGTCTTACC | CCCCCTGGCCC | GCCCAATATA | CGGTGTGAC   | GTCTGATGTA  | CAGGCCCGGG  | 8580  |
| CGTCGCGGGC | GTTGGTGGGA  | AAACCGGCAC | CACCCTGTGC  | GSCCGAATCC  | GCCACGGGGG  | 8640  |
| CTGCCAGACA | GTACACTGTC  | TCCAGCAGCG | ACTTCAGTCT  | CTTGTGACTT  | TGGGCGCTCA  | 8700  |
| CCACCAAAAA | TTGCAAAACC  | TGCCTGTAGT | CCGTGAAGTA  | GGTACGGCAT  | ATTACCATGG  | 8760  |
| AGTTGTACAC | GCCCAGGTTT  | TTTGAGAACA | CCAGGCTCGC  | CTTGAACTTT  | GTAAAGTCAT  | 8820  |
| CCTGCCCCAG | CACGACAGAC  | GTATTTTTTG | CAAGGTATAC  | GTCCGACTCC  | ACGGGAAGGA  | 8880  |
| CGTGCCCCAA | CTGGGACACG  | GCGTCGCTTG | GTCGGCACAG  | AAAGCACTTC  | AGGGTTGTGG  | 8940  |
| AAAGGCCATT | ATTCGATATA  | ACAAAGCAGG | GAGAGAACGG  | GTAGTGATC   | TCCTCCAGGA  | 9000  |
| GGTGCGCCCA | AAACTTATAC  | ACAAACTCTA | AGTGGTACAC  | GCAACCGTGC  | TGCATTCTAA  | 9060  |
| CCGTACATAT | GGCGGTAGCA  | CCGCCCTTAG | CATAAACTGG  | GGCCCCGTCC  | ATGCACCGTT  | 9120  |
| CCAAATCCAG | GGACTGACCA  | GACTGTCCCA | AGTATGAGGA  | TACCACCCGA  | CACAGTTGGT  | 9180  |
| CCACTACACG | CTTACCAACG  | ACACTCATGG | CGACAGCGGG  | GTGGGGCTGG  | CAAGGCCCCC  | 9240  |
| AAAGCGCGAC | ACCCGCAGTC  | AATCAGGGCC | GTGCCCGCGC  | CTCGGAGAAT  | ACGGCGTCCG  | 9300  |
| TGCTCAGGAT | CTTGCGCAGG  | ACCTGCCTTA | CCGTGTCCAC  | CTTGCTCTCC  | AACACCAGAG  | 9360  |
| TATGATCGCA | GGCTGCAGGC  | TGTGCCCGCT | GGACGAGAAA  | GGTTTTTAAA  | TACTGACAGT  | 9420  |
| AGTTGATGGC | GTTCAATCTA  | CAATAGATCG | TGGGAATAAA  | AATTTGCATG  | TCACGAGGCA  | 9480  |
| GAAGCTGGTC | AGACGGGTAC  | TCCATGTTGG | GTTCCACGGG  | GAGGGGAACA  | CACGCCCCAA  | 9540  |
| GACACGACGG | CGCACATAGG  | GAGCGGAGCA | AACAATTGAT  | TCAAATATTT  | GACTCCGCGAG | 9600  |
| CGAGCCGGTT | TGCAGAGTGG  | TCACCTGCCC | TGCTCCACAC  | CCACCCCGGC  | GTCTCTTCCA  | 9660  |
| ACTCTCAACT | CACGATCCAG  | GGAAACCACC | GTCCAGTGGC  | CATGTTTGGT  | CCCTGGCAAC  | 9720  |
| TGGGTACAAT | TACCCGTCCAC | CGAGATGAGC | TCCAAAAACT  | ACTGGCAGCC  | TCCCTGCTCC  | 9780  |
| CGGAGCACCC | GGAGGAGAGC  | CTCGGTAAAC | CCATAATGAC  | ACAGATTCCAC | CAGTCGCTCC  | 9840  |
| AACCATCTTC | CCCCTGACGG  | GTCTGTGAGC | TCCTATTTTC  | TCTGGTCCGC  | GATTCTGTCCA | 9900  |
| CCCCCATGGG | TTTCTTCGAG  | GACTATGCCT | GCCTCTGCTT  | CTTCTGTCTA  | TACGCCCCAC  | 9960  |
| ACTGCTGGAC | CTCGACCATG  | GCGGCAGCGG | CAGACCTGTG  | CGAGATCATG  | CATCTGCACT  | 10020 |
| TTCCAGAAGA | GGAGGCGACA  | TACGGGCTAT | TGGGACCGGG  | TGGCCTTATG  | GGTATCGACT  | 10080 |
| TGCAGCTGCA | CTTCTTTGTT  | CAAAAGTGCT | TTAAGACCAAC | CGCGCGCGAA  | AAAATACTGG  | 10140 |
| GAATATCCAA | CCTGCAATTT  | TTAAATCAG  | AATTCATCCG  | GGGCATGCTT  | ACAGGCACCA  | 10200 |
| TCACCTGCAA | CTTCTGCTTC  | AAAACGTCTT | GGCCGAGGAC  | AGACAAGGAG  | GAGGCCACCG  | 10260 |
| GCCCCACCCC | ATGCTGCCAG  | ATTACAGACA | CCACCACCGC  | ACCCGCGAGC  | GGCATACCGG  | 10320 |
| AACTAGCCCC | GGCCACATTC  | TGCGGCGCAA | GTCGCCCCAC  | AAAGCCCGAG  | CTACTTCCCC  | 10380 |

|             |            |             |             |            |             |       |
|-------------|------------|-------------|-------------|------------|-------------|-------|
| CGCTAATAGA  | TATCTGGTCC | ACGAGCTCAG  | AGCTCCTTGA  | CGAGCCCGCG | CCTCGACTGA  | 10440 |
| TCGCAAGCGA  | CATGAGTGAA | CTCAAATCCG  | TGSTCGCATC  | CCACGATCCG | TTCTTCTCTC  | 10500 |
| CCCCGCTTCA  | GGCAGACACC | TCACAGGGTC  | CATGTCTGAT  | GCACCCAAAC | CTGGSGTTAC  | 10560 |
| GATACAAAAA  | CGGGACTGCA | TCCGTCTGCC  | TCCTCTGCGA  | GTGCCTTGCG | GCACACCCAG  | 10620 |
| AGGCACCCAA  | GGCGCTGCAG | ACCCTTCAGT  | GCGAGGTAAT  | GGGCCATATA | GAAACCAACG  | 10680 |
| TAAAGCTGCT  | AGACAGAATT | GCCTTTGTGT  | TGGACAACCC  | ATTGCGCATG | CCATATGTAT  | 10740 |
| CAGATCCGCT  | ACTTAGAGAG | CTGATCCGGG  | GCTGTACCCC  | ACAGGAAATT | CACAAGCACC  | 10800 |
| TGTTCTGCGA  | CCCCTGTGTC | GCCCTCAATG  | CTAAGGTGGT  | GTCAGAGGAC | GTACTATTCC  | 10860 |
| GCCTGCCCAG  | GGAGCAGGAG | TATAAAAACC  | TCAGGGCATC  | CGCGGCCGCG | GGACAGCTCC  | 10920 |
| TCGATGCCAA  | CACCTGTTC  | GA CTGCGAGG | TCGTGCAGAC  | TTTGGTCTTT | CTCTTTAAGG  | 10980 |
| GTCTCCAAAA  | CGCCAGGGTG | GGGAAAACCA  | CCTCACTAGA  | CATTATTCCG | GAGCTAACCG  | 11040 |
| CACAACTAAA  | AAGACACCGC | CTAGACCTGG  | CCCACCCCTC  | ACAGACGTCA | CAC TTGTACG | 11100 |
| CTTGAGCTGG  | TCCCGGGCCT | TCGCACCCCA  | TCCACCGATG  | CCGAAATCAG | TGTCCAGCCA  | 11160 |
| CATCAGCTTG  | GCGACCTCAA | CCGGTCGCAG  | TGGACCGCGA  | GACATCAGAA | GATGCTTGTC  | 11220 |
| ATCCCGCCTG  | CGSTCGGTCC | CGCCCCGGGC  | GCGAAGCGCC  | AGCGTCAGCA | GCAAGCAGAG  | 11280 |
| AAACGGCCTT  | CGCAAGTTTA | TCTCAGACAA  | GGTATTTTTT  | AGCATCCTAT | CGCAGAGACA  | 11340 |
| CGAGCTAGGA  | GTGGACTTTT | TCCGTGAGAT  | GGAGACCCCG  | ATATGCACCT | CCAAAACAST  | 11400 |
| AATGCTGCCC  | CTAGACCTGT | CTACCGTCGC  | ACCCGGCCGC  | TGCGTCTCCC | TCTCTCGGTT  | 11460 |
| TGGACACTCC  | TCAAACATGG | GGTTCCAGTG  | CGCTCTGTGC  | CCATCCACAG | AAAATCCCAC  | 11520 |
| CGTTGCCCAA  | GGCTCCCGGC | CTCAGACAAT  | GGTGGGCGAT  | GCGCTCAAAA | AAAATAACGA  | 11580 |
| GCTATGCTGG  | GTAGCGCTGG | CCTTTTATCA  | CCAGCGAGAC  | AAAGTGATCC | AACACAAGAC  | 11640 |
| GTTTTACSTA  | TCACTCCTCA | GTCACTCCAT  | GGATGTGGTT  | CGGCAGAGCT | TCCTGCAGCC  | 11700 |
| TGGTCTACTG  | TACGCTAACC | TGSTCCTAAA  | AACCTTTGGG  | CACGATCCCC | TACCCATCTT  | 11760 |
| CAC TACCAAC | AACGGCATGC | TAACAATGTG  | CATCCTTTTT  | AAAACCCGGG | CAC TACATCT | 11820 |
| GGGAGAAACT  | GCGCTTAGGC | TGCTTATGGA  | TAACCTCCCC  | AACTACAAGA | TATCGGCGGA  | 11880 |
| CTGCTGCAGA  | CAGTCCTACG | TGGTCAAGTT  | TGTCCCAACG  | CACCCGGACA | CCGCAAGCAT  | 11940 |
| TGCAGTGCAG  | GTACACACCA | TATGCGAAAC  | GGTTGCGGCG  | CTAGACTGCA | CCGACGAGAT  | 12000 |
| GCGGGATGAC  | ATTCAAAAGG | GAACCGCACT  | TGTCAACGCG  | CTATAACCTC | ACATGTAGCC  | 12060 |
| TGTCACCCCA  | GCTCCTATTG | CAACTGACCA  | TGTTCAAGGTG | GTAATAAAGT | CATTAAACGA  | 12120 |
| CAAAAGTGATT | CTTTTAATCT | GTTTATTGTT  | TTTGAACATG  | TGGCACACGC | TGCAATGTAC  | 12180 |
| TGCCATGAAA  | GCTGTTTCTA | TATCCACCAC  | TTGGCGTCTG  | CCGAAGTCAG | TGCCACAATT  | 12240 |
| TCATTAACAA  | ACAAGGTCAA | TACATTGTGA  | GGGAGTGTGT  | TTTGCCATGG | TACCATTCGT  | 12300 |
| GTGGTTTGGG  | AGAGCGGACG | CCATTTGCGT  | GCAAAATGTG  | CTTTGCTGGA | GSCCAACTTC  | 12360 |
| CGTCGCGCTG  | GTTGATGCGC | GGCACATTGT  | GTCAAACGAG  | GCACCCCTCC | CCACCGAGTG  | 12420 |

|            |            |            |            |             |             |       |
|------------|------------|------------|------------|-------------|-------------|-------|
| CTTTAATGCG | GAGAGGAATG | GTGGCCTGGT | TGACACCGCG | TGCCGGCCAT  | CTGAACTGTG  | 12480 |
| ACTGTGTTAT | GAGCCACGGG | TATGCCCTCG | ATACGCCTGC | TCTTCAGCAT  | TGTATGTGTT  | 12540 |
| TAATGTTGTG | CTTGGTGCAA | CCGTGATTGT | GTTTTTGTAT | TTTATTTTAC  | TGACACTCTT  | 12600 |
| TGGGAGGGCA | CGCTAGCTTC | AGTGCGCGCC | CGTTGCAACT | CGTGTCTCTGA | ATGCTACGGG  | 12660 |
| GCCACGCTGG | CCACTCGGGG | GGACAACACT | AATCGCCAAC | AGACAAACGA  | GTGGTGGTAT  | 12720 |
| CGCCCCAAGC | CTCCAGCGCC | ACCCATTTAG | TAACACATCC | GGGACATGAA  | CTGCCACAAA  | 12780 |
| CACCGTTAAG | CCTCTATCCA | TGCATTGGGA | TTGGAGTGAG | GAGGGAGGAG  | GGCACCAGGT  | 12840 |
| TCCCCGGGAG | GAGGGCACCA | GSTTCCCGGG | GAGGAGGGCA | CCAGSTTCCC  | GGGGAGGAGG  | 12900 |
| GCACCAGGTT | CCCCGGGAGG | AGGGCACCAG | GTTCCCGGGG | AGGAGGGCAC  | CAGGTTCCCC  | 12960 |
| GGGAGGAGGG | CACCAGGTTT | CCGGGGAGGA | GGGCACCAGG | TTCCCGGGGA  | GGAGGGCACC  | 13020 |
| AGGTTCCCCG | GGAGGAGGGC | ACCAGGTTCC | CGGGGAGGAG | GGCACCAGGT  | TCCCGGGGAG  | 13080 |
| GAGGGCACCA | GGTTCCTGGG | GAGGAGGGCA | CCAGGTTCCC | GGGGAGGAGG  | CTGGGGTGCG  | 13140 |
| CCGCGCCGGG | TTCTTGGGGT | GCGCCGCGCC | GGGTTCTCTG | GCTGCGCCGC  | GCCGGGTTCC  | 13200 |
| TGGGGTGCGC | CGCGCCGGGT | TCTTGGGGTG | CGCGCGCCGC | GGTTCTCTGG  | GTGCGCCCGC  | 13260 |
| CCGGGTTCTT | GGGGTGCGCC | GCGCCGGGTT | CCTGGGGTGC | GGGGTGCGGG  | GGACCGCGCC  | 13320 |
| GGGSTACTGC | AGGGTTCGCA | GGGTTCTGGG | GTACTACCTG | GTTTCTCTGG  | GTGTGCCAGG  | 13380 |
| ACGGGTTCTT | GGGGTGCCAC | CGCTCTCTGA | TACGTGTAAA | TCCAAGAGAT  | CCGTCTCTCC  | 13440 |
| TSCCGCCGCG | CGCGTAATGC | GCGAGGGGGG | TGGGTCTCCC | CTGTTCTTTA  | TAGCGTTTCC  | 13500 |
| TGCSAAGGGG | GCGTAACCGT | AGGACAAACT | GCTTATGTAG | GGGTTAGCCA  | CCCATTTCCC  | 13560 |
| GGGGCCGCGC | CAGAGGTGAG | CGTGGACTTA | GCATCCCGCT | CCCATTTACC  | GAAACCCACC  | 13620 |
| AGAGGGCGAG | TTCCAGGGCC | GTGACTCACT | AGCTCCCTTC | CCATCGAACA  | ACCAGGCTTG  | 13680 |
| GCTAACACGG | CTGGAGTGCC | GGTGGGCGGG | GCCCCATATA | TCTTGGCCCC  | CATCTACTGA  | 13740 |
| AACGACCCAG | TAGAAAAATC | CCAACCCCAT | GACTCATCAG | GCCCCATTAT  | ATAGAATATC  | 13800 |
| CCASTAGAGT | GACCCAGCTG | GTTTCCATAA | ATGGATATAC | TTCCGGAAAA  | CGAAGGAGGG  | 13860 |
| TTGAATACAG | TTGGGGGTAG | TCCGCTGGTA | TTCCAGCTG  | AGGTTGCCTT  | ATTTGGTAAT  | 13920 |
| GCTTCCGGAA | ATACCACCTG | AGTACCCCAT | TGSTTTATAC | CTTGTTTAAAT | TGTAGAATTA  | 13980 |
| CAGCTGGATT | TACCCAGCCG | GSTTTACGCA | GCTGCGTATA | CCCAGCTGTG  | TTTACGCAGC  | 14040 |
| GGGGTTTTAG | CAGCTGGGTA | GACCCAGCTG | GSTATACCTA | CTGGAATAGG  | GGGTGGGATG  | 14100 |
| ACTCAGCTGC | GCTAGGATTA | AAGGATTATA | TATATATATA | TAGGAAAAAT  | CAAAACAAAA  | 14160 |
| CTCTAATCGC | TGATTGGTTT | CCGCTCTGGG | CCAATCAGCT | TGGGAGTTCT  | AGGGATAGGG  | 14220 |
| GCCAATGGGA | GGCTCCGAA  | TTTGATTGAC | GGCTGGGGCG | TCCAATGGAA  | TGGGCGGGTC  | 14280 |
| GCCTAGCTGC | AACGGGATTG | GTGGGCGGGA | TGGGCGCAAT | GCGGCTCGGA  | AAACTTTGAT  | 14340 |
| TGACGGGGCG | GCGGACCAAT | GGGAGCGGGG | CAGAGGATTA | TGGGGGATTA  | GCAAAATTCAA | 14400 |
| GATGGCGGGG | CCCATGAAAT | GGCCAAAAAT | TATAATTTTT | CGAGTCGGTC  | ACGSTCCCAT  | 14460 |

|             |            |            |            |             |             |       |
|-------------|------------|------------|------------|-------------|-------------|-------|
| CTAGCGGCGT  | GACCTGGAGG | TGACCCCGTG | CACCCGGGCG | CTCTGAATTT  | TTCTGCGCAT  | 14520 |
| GCGCGACTCC  | TCATCTACAT | AATTTATGCA | CATAAAAGGA | TTAGCGCATG  | CAAATTAGTC  | 14560 |
| AGATAGCAGG  | GCCATCCACA | CTTTATGTTG | GCCGCGTGCC | AGGCGCCGGC  | GTGGGCGCGG  | 14600 |
| CGCGCGTGCT  | CTCTCAGTCG | CGCCTAGCTG | CTTCCAACAG | ACAAAAGCGG  | GGCGTTAGTG  | 14700 |
| AGGGAGTGCG  | CGCGCTGCGT | TGACTTGGCC | GATTTCCAGT | GCATGCTTTG  | TCACCCGAGC  | 14760 |
| GCGAGAATGG  | AATTTTCATT | ATTGAGCAAT | TTGGGCACCC | TGGGCACGAT  | AACCATACAT  | 14820 |
| GGATACACGG  | GTTCCAAATA | TGCAAAGTAG | ACACTAAGGT | ACCATTITGGC | ATATTTGGAC  | 14880 |
| GTCTTGGGCA  | GGTTAGCTAC | CCACCAGAAT | ATATGGGACT | CTGGGCAGGA  | TAGCCACCCA  | 14940 |
| CAATTGTTTT  | GCGCCCTCT  | TTGGCCAGGG | GACCAAGGTC | GTATGSTTCG  | CGCTACACTA  | 15000 |
| AGCCCGAAGC  | TTGAGCTTTG | CGTGCTTTTC | ACGTCCAGGC | GGCTGGCACA  | CGGGCCGTGA  | 15060 |
| GCGCCAGCAA  | CATGGGATCA | TGGTAGTAAG | ATACAGCATA | AATCCCCGTC  | CGGTGGCGCT  | 15120 |
| CAACGCCAAT  | ATGCGCGGCT | GCGTGGTATC | TCATCGGTGG | GCACGCGTAC  | GGTGGTCTCA  | 15180 |
| TGGGTATTGG  | ACTTGTAGGC | GAGGGGAGGC | GCATACGACA | AAAATTGCCC  | CCGTGAAGGT  | 15240 |
| CGGGAACCCG  | CCCGCGCTTC | CGCAAGGCAC | GGGGCCGCAT | CGGACACAGG  | CTAAGCATTG  | 15300 |
| AGGATCATAA  | CACCGCCCTA | GAAATGTTTA | AGCTGTGACC | AAAGCGAACC  | TCCCATGAGG  | 15360 |
| CATACGCGAG  | CGTGGAGGTA | GGATTCCCAA | GGCTATTGAG | AGACGSGTGG  | TGAAATGATG  | 15420 |
| AAGAACACAC  | AGAACAATAA | CGGGCGACTA | GATAAAAAGA | CTCGCTCAAC  | AGCCCGAAAA  | 15480 |
| CCATCAGCCC  | GACCGCCGAT | GGATTAGGTG | CTGCTGGACA | AGTCTTTCTA  | AACCCGCGCA  | 15540 |
| GGSTTTGTGT  | CGATCCAGAC | GCTTACGAAC | GCCCCGTTTA | AAAACACTAT  | TCATAATTAA  | 15600 |
| CAGAAGTTGA  | CACCAGCCCG | CAGTTACCCA | ACCTTCTATT | TTTTTGGAGT  | GTTGACAAGT  | 15660 |
| TTCCATCGCC  | CGTTTGGCGT | TTCCCGCATG | GTGTCAAAAT | AGTGACGCAC  | CCTCCCCCGG  | 15720 |
| TCACTATGGG  | TTTACCCTGA | TTTAGTAAGT | AAAAGTGCCG | CCCCCGCCCA  | CTCATTTTTT  | 15780 |
| TACCCGTGTA  | TTTGCTGTAT | TTACATCTAC | GGACCCCTTT | TTGGTGAGAT  | TGCCGTGGTT  | 15840 |
| CTAAATAAGG  | TTGTGGTTTT | CGGACCCCTT | CAGGGACCAA | ATCTTTTACG  | TGTTGCCAAG  | 15900 |
| GTAGCATTTG  | CTGGACCCGT | ATAGGTTTTT | GTGGCACCAG | GTTATGCTCT  | TATGAGCGGG  | 15960 |
| CTTGACCGGC  | AAGTTCCAGG | CATCCTAAGT | GCTTGATGTA | GACCCCTTAGG | GCACCAGGGA  | 16020 |
| CTACCTAGGT  | CAAACTCCCG | CTTAGTCATG | ACGCGGTGCG | CACGAGSTTT  | GAGAGGCGTA  | 16080 |
| GACATCCGTG  | TCGACTGCTG | GACGGAGGTA | GTATAATCAG | CTAGGCCTCA  | GTATTCATATG | 16140 |
| TACAAATGA   | ATGCCCTAGA | GTACTGCGGT | TTAGCTAGTT | ATACTGCCCG  | GTTCCACCAG  | 16200 |
| GCGGCGTTGT  | GGCCACGGGC | GSTTCGTGCG | TTGGACCTGG | AGGGGTGTCA  | CATTCTGTGA  | 16260 |
| CCGCGACGTT  | GACGTTAGAC | ACACGTCGCT | GCGCTCCTCA | GAATGTGATA  | GCCCATCACA  | 16320 |
| GGCATTTGTAG | CTGTTGCGTT | GSTTGGGAGT | TTGGGGACCA | AATTTCTATA  | ATTGGTGTCA  | 16380 |
| CCGCGGCAGC  | TCTAGCCCTG | GAAGATCTGG | AAGCTTGCTT | CAATGCTCTA  | GATCGACCCG  | 16440 |
| GACTACAGTT  | AGCGAAGTAG | ACCCATTATA | ATCTTAATCT | TAAATCTGST  | TGACGGACTT  | 16500 |



|             |             |             |            |            |             |       |
|-------------|-------------|-------------|------------|------------|-------------|-------|
| TCGCGCCGGG  | AACACGCAGG  | TGGCAGCGGA  | TGTGTTTTGC | CCAAACACGA | GGGTTGCAGG  | 16560 |
| AAACAGGTGC  | TGCCGGGGAT  | TATGTACAGC  | TTACACCCAG | TTTCCTGTAA | TGCCCCGCAT  | 16620 |
| CCGCGCGTCC  | TGGGCAGCAC  | CGCACCCCTGC | GTAAACAACC | GGGTACTTTT | TCCTCCCTCCC | 16680 |
| CCCACCCCCA  | CATCCTTCCT  | CCCACCCCTGC | CAGTCCAACC | CGCTTCCTGT | TTTATTGCGC  | 16740 |
| TTCAAACAGA  | AGCAGGCATT  | CTAATGATTC  | TTACAAAAC  | TGTTAGTGTT | TATTAAATCA  | 16800 |
| GATACATACA  | TTCTACGGAC  | CAAAAATTAG  | CAACAGCTTG | TTATCTATGG | TGTATGGCGA  | 16860 |
| TAGTGTGGG   | AGTGTGATGG  | GCCGGAAAGG  | TGAAGSCCCA | TTAGGGTTTG | CACTTGGCGC  | 16920 |
| TGTAGGTCTA  | CTCTTGACAA  | AGATCTAAGC  | ATTGACATTA | GGGCATCCAC | GTCAGTGGGA  | 16980 |
| CCCAGTAGGT  | CTAAGTTTTC  | CATACAGTAC  | ACCCAGTGTA | AGATGTCTGT | GGTGTGCTGC  | 17040 |
| GAGACCCAT   | AGTGTCTTG   | CTTAAAAATA  | TCAAAGACCT | AATATCCCTC | GCACACAGCT  | 17100 |
| CCCCGTCTAC  | GTGGAGAACA  | GTGAGCTGAT  | AAGGGCTGAA | ATAACTCATT | GTGCCCCGTA  | 17160 |
| GGTGGCGCTC  | TAAAAACGC   | GGGTCTAAGT  | GAAGCAGGTC | GCGCAAGAGG | TCTCTGCGAC  | 17220 |
| CTGCACGAAA  | CAGACATTCC  | GCTAACAGGG  | GAAACGTTAA | CCTGCCCTCC | TCCTTTAAAG  | 17280 |
| CTCTAAGAGC  | TCCAATTAA   | TGGGCCAGTG  | TGGGTTGAGG | TATGAACACG | TTTAGGAGGA  | 17340 |
| ACAATACCAC  | TTCCCTGTCA  | TCCGTGCCCA  | GTTTCCGCGC | CACCTCACAG | AGAACCTCGT  | 17400 |
| AAGTGGCCAT  | GGTGCCGCT   | TGTATATGTG  | AAGGCACCGA | TGTGGAAAAA | CAAAGGAAAA  | 17460 |
| TTTATTTTTT  | CGCCCTAAAC  | AAAATCACAA  | GCTTAATAGC | TGTCCAGAAT | GCGCAGATCA  | 17520 |
| AAGTCCGAAA  | CAGATGTTAG  | GATCTGTTC   | ACTGCCGCT  | GTAAGACGGA | AACATCGCAT  | 17580 |
| CCCAATATGC  | TTGCCAGCTG  | AGGAACATCC  | CCACCCGAGT | GGGTATCCTG | CGGAATGACG  | 17640 |
| TTGSCAGGAA  | CCAACAGCGC  | ACAGCCCTGCA | GCGCTGATAA | TAGAGGCGGG | CAATGAGCCA  | 17700 |
| GTCTTTGGST  | CAACTAAGGC  | TTTTGTAATC  | AGGGTGTGTA | CCTCGTGSTG | CCAAAAGTCC  | 17760 |
| AGGTGTGGG   | AGCCCCCCAG  | CAATTTAAGT  | AACAAGAAGG | AAGTGACGTC | CGTCGCTAAG  | 17820 |
| ACTGCCCTCTG | TTGCCACGCG  | CAACTTCTCA  | AGGAGTTCTT | TCTCCTGSTC | TATAAGTTCT  | 17880 |
| TGGCGGGAAA  | AGGAGTCTGC  | CGCGGCATAG  | CAAAGTGAAC | TGTTAGAAAT | AGGCGTGAGG  | 17940 |
| CTTCTGAGCT  | TACTGGCCAC  | TAACAGGCAG  | GCGCTCCCTG | TCTTTTGAAA | GTGTTCTTTG  | 18000 |
| GACACCTGCT  | TTATAAGTAG  | GAGTGTGTCC  | AAAAGATTAA | GGGCCAACGC | GACCACGTTA  | 18060 |
| GGTTCTAGST  | TGTATTCTCTG | GCAAACTGAA  | AACATCCATG | TGCCCAGTAA | CTTACGCATA  | 18120 |
| TGCGAAGTAA  | GAGATTGTTG  | AAAGGTCCCA  | AATACAGAGT | CAGAAGTTAA | AAAGCGCGGC  | 18180 |
| TCAATTTCAA  | GAATATTGTA  | AAAGATCCGA  | TCCTCACATA | GCGTGGGATC | CAGAAGTCCC  | 18240 |
| GAGGGCGGGT  | TATTGGCAGT  | TGCCATATAG  | AGTGGCGAGC | GTATGTGGCC | TACCTGTAGA  | 18300 |
| GCCTGGAGTT  | TCAGGGTGGT  | CTGTGAGGTT  | CTCCCATCGA | CGACGCTGGG | CGCGGAGAGT  | 18360 |
| ACGCTAGCCG  | TTGTCCCTGT  | GTTCAGTTGA  | GSTAGATGGG | TGCTGAGAAC | ACTGCCCCCC  | 18420 |
| ACACACACCA  | GCACCCATGG  | CGCCAAATGC  | AAGTGCGGAG | CGGCGACGGT | GGCTTCTAGG  | 18480 |
| GAGGAAAAAG  | GGGGAGAGGT  | GTGGCTTTTA  | TGTCATTTCC | TGTGGAGAGT | CCCCAGGACC  | 18540 |

|            |            |            |            |            |            |       |
|------------|------------|------------|------------|------------|------------|-------|
| TTGGTTTTCC | CCTGGCTGGG | TTAATGGCAG | GGGCTTTTTA | AACTTAACTA | TGGAAGATTG | 18600 |
| TAGGTTTTCC | GCCAGGGGGT | GACTAGCTTC | CCAGGCTAGG | CGGGCCATTT | GTACTTTCTT | 18660 |
| ACTTGTGTCT | TTGTTCTGAC | AATACACATA | TACACAATAA | GTTATGGGGC | ACTGGTCTGG | 18720 |
| TCCAGGGTGG | GGCAAGCAGG | ACACGGGGCC | TGCCTTTACT | CCTCCAAACT | GGAAGGCGTG | 18780 |
| AGATAATTTT | TAAAGTCCGT | ATGGGTCAAT | GCCCCAAAAA | ATCACTGCAA | ACTTCCATTC | 18840 |
| ACACTTTTGA | TCTCTCTTTC | CATCCTTTCC | CAAAAAGCGT | CTATAAAAGA | TGTCTTCTGG | 18900 |
| CCTAGCTTTC | GCAGGACAAT | CATCTATCTG | TCTGTAAGGG | ACCGGTGGTT | GTTGGTATCT | 18960 |
| TGGATGTGGC | TTTTTTGGGT | GGGTAACTGG | AACGCGCCTC | ATACGAACTC | CAGGTCTGTG | 19020 |
| GGGTGGTGAT | GTCTGAGTA  | CATAGCGGTA | TTGCGGAGAT | GGGCCAGGTT | GTGGGTCACT | 19080 |
| GTCTGGTGTA | TTATCTCCTG | GTGGGCTACT | GGCAATTTGT | TCATGTGTGC | TAACAACAGG | 19140 |
| GTAATCCACT | TCCATTTTCT | CCTCGGATGA | CGACCCCTGC | AAGATTATGG | GCTCTTCCAC | 19200 |
| CGTCTCTGTC | TCCTGCTGTT | CCACCCCTCT | CTGCTCCTGC | TCTTCCACCT | CCTCTAACTC | 19260 |
| CTGCTGCTCC | TGCTCTTCCA | CCTCCTCTAA | CTCCTGCTCT | TCTGCTCTTT | CCACCTCCTC | 19320 |
| TAACTCCTGC | TCTTCTCTGT | CTTCCACCTC | CTCTAACTCC | TGCTCCTCCT | GCTCCTCCTG | 19380 |
| CTCCTGCTCT | TGCTCCTCCA | CCTCCTCTAA | TTCTGCTCTT | TCTGCTCCTT | GCTCTTGCTC | 19440 |
| TTCCACCTCC | TGCTCTTGGT | CTTCCACCTC | CTGCTCCTCT | AACTCCTGCT | CCTGCTCCTC | 19500 |
| TAACTCCTGC | TCTGCTCCTT | CTAACTCCTG | CTCCTGCTCC | TCTAACTCCT | GCTCCTGCTC | 19560 |
| CTCTAACTCC | TGCTCCTGCT | CCTCTAACTC | CTGCTCCTGC | TCTCTAACTT | CCTGCTCCTG | 19620 |
| CTCCTCTAAC | TCTGCTCCTT | GCTCCTCTAA | CTCCTGCTCC | TGATCCTCTA | ACTCCTGCTC | 19680 |
| CTGCTCCTCT | AACTCCTGCT | CCTGCTCCTC | CTGCTGCTCC | TGCTCCTCCT | GCTGCTCCTG | 19740 |
| TTCACTCCTC | TGCTGCTGCT | CATCCTGCTG | CTGCTGCTCA | TCTGCTGCTT | GCTGCTCCTC | 19800 |
| CTGCTCCTGC | TGCTCATCCT | GCTGCTGCTG | CTCATCCTGC | TGCTGCTCAT | CCTGCTGCTC | 19860 |
| CTGCTCATCC | TGCTGCTCCT | GCTCATCCTG | CTGCTCCTGC | TCATCCTGCT | GCTGCTCATC | 19920 |
| CTGCTGCTGC | TCATCCTGCT | GCTGCTCATC | CTGCTGCTGC | TCATCCTGCT | GCTGCTCATC | 19980 |
| CTGCTGCTGC | TCATCCTGCT | GCTGCTCATC | CTGCTGCTGC | TCATCCTGCT | GCTGCTCATC | 20040 |
| CTGCTGCTGC | TCATCCTGCT | GCTGCTCATC | CTGCTGCTGC | TCATCCTGCT | GCTGCTCATC | 20100 |
| CTGCTGCTGC | TCATCCTGCT | GCTGCTCATC | CTGCTGCTGC | TCATCCTGCT | GCTGCTCATC | 20160 |
| CTGCTGCTGC | TCATCCTGCT | GCTGCTCATC | CTGCTGCTGC | TCATCCTGCT | GCTGCTCATC | 20220 |
| CCGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | CCGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | 20280 |
| CCGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | CCGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | 20340 |
| CCGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | CTGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | 20400 |
| CTGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | CTGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | 20460 |
| CTGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | CTGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | 20520 |
| CTGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | CTGCTGCTGT | GGCTCCCGCT | GCTGTGGCTC | 20580 |

|            |             |            |             |             |             |       |
|------------|-------------|------------|-------------|-------------|-------------|-------|
| CTGTGGCTCC | TGCTGTTGTG  | GCTCCTGCAG | GGGCTCCTGC  | TGCTGTGGCT  | CCTGCTGCTG  | 20640 |
| TGGCTCCTGC | TGTTGTGGCT  | CCTGCTGCTG | TTGTGAACCT  | TGGATGCTCA  | ACGTTTTGTT  | 20700 |
| TGCATCGCCC | CCGTCCCTCT  | CGTCCCTCCT | CTTGTCTCTC  | TCCTCGTCAT  | CCTGCTGCTG  | 20760 |
| CTCATGTGTC | TCATCATCGT  | CATCCTCCTC | GTCTCTCTCT  | TCCTCGTCCT  | CCTGCTGCTG  | 20820 |
| CTCTCTCTCT | TCCTCTCTCT  | CGTCATCCTC | CTCGTCATCC  | TCCTCGTCAT  | CCTGCTGCTG  | 20880 |
| ATCCTCTCTG | TCATCCTCCT  | CGTCATCCTC | CTCGTCATCC  | TCCTCGTCAT  | CCTGCTGCTG  | 20940 |
| ATCCTCTCTG | TCATCCTCCT  | CGTCATCCTC | CTCGTCCTCC  | TCATCTGTCT  | CCTGCTCTCT  | 21000 |
| CTCATCATCC | TTATTGTTCAT | TGTCATCCTT | GTCAACCTGA  | CTTTCTCTTG  | TAATCTCTGT  | 21060 |
| GTCCCCATTA | TCCTCGCCAG  | CCTGATTATT | TTCCGGAACAT | TCTTTTTTCAT | TCTTGGATGC  | 21120 |
| TTCTTTCTGA | ATCTCCGCAA  | GGAGCACCAA | CATGGCTGTG  | TCATCACCCC  | AGGATCCCTC  | 21180 |
| AGACGGGGAT | GATGATCCTA  | TGGAGATGGG | AGATGTAGGC  | GGTTGGCGTG  | GCGGASTATC  | 21240 |
| GCCATCGCTG | GATGATCCCA  | CGTAGATCGG | GGACTCTGTG  | GCCCATGGGG  | GGTACACACT  | 21300 |
| ACGTTTGGCG | AAGTCACATC  | TAGGGGGAGA | GACTGGGGGC  | GACTGACATA  | TTGGGTTTAG  | 21360 |
| TGTAGAGGGA | CCTTGGGGGG  | ACGATAGCCT | TCTTTTTCTC  | AGGCTACGCA  | GGGTAGACGG  | 21420 |
| AGCTAAAGAG | TCTGGTGACG  | ACTTGGAGGG | AGGCTCGGGT  | GGAGGAGTCG  | TGGGTGAGTG  | 21480 |
| TGGAGGTGTA | GTCTGCTGCG  | AGGGTGGCGG | ACGCATAGGT  | GTTGAAGAGT  | CTGGCCTTCC  | 21540 |
| TGTAGGACTT | GAAAGCGGTG  | GCCTTTGAGA | AGACTCTGGA  | GACTGCGTGG  | GTGGCAATGC  | 21600 |
| AGGAGATGGA | GAATGAGTAT  | CCGTGGTCCC | CGGAGACACA  | GGATGGGATG  | GAGGGATTGG  | 21660 |
| GGAGGAAGAG | GTGGTTACGG  | GGGGTAAGAG | TGCTGGTGGG  | GGTAAAGGTG  | TTGGGGGAGC  | 21720 |
| GGGTGAAGGA | ATGGGAGCCA  | CCGGTAAAGT | AGGACTAGAC  | ACAAATGCTG  | GCAGCCCGGA  | 21780 |
| TGTGAACACT | GTGGGACTTC  | CAGGTATAGG | CAAGGTGTGG  | GGTCCACATT  | CCCGGCCGTC  | 21840 |
| GATGGAGTGG | GCGACATGCT  | TCCTTCGCGG | TTGTAGATGT  | AGGTCATCGC  | CAAGGTGACA  | 21900 |
| TCTTTCCGGA | GACCTGTTTC  | GTTTCCTACA | ACTTCCTCTC  | GTTAAGGGCG  | CGCCGGTCTT  | 21960 |
| CCGTCCCGAC | CTCAGGGCGA  | TTCCCGGGGG | CGCCATCCTC  | GGGAAATCTG  | GTCTGACAAAC | 22020 |
| CAAAGTAAAA | TTATGGAGGC  | GGTGGCAGTA | TATTCACATT  | ATGCAATACC  | CGTAGTGACC  | 22080 |
| ACAAGSGGGA | GCTCTCAGAC  | AATTAAGCGG | TTACACACAG  | TAGCAGGCTG  | CAGTACCGCC  | 22140 |
| CATGGCCACA | GGATGTAGAT  | CGCAGACACT | GAAACGCTGA  | AACACAGCAT  | TAAGCTGCAA  | 22200 |
| TACCGCCGAT | GGCCACCAGA  | TGGCACGCGC | CGCCAGCAAA  | TTTAAGTCTT  | GGTGGCTCAC  | 22260 |
| CTGCCAGSTA | AACAAGGTTA  | AAGTGGGTTT | GCTGGCCTTG  | CGTTGCCATG  | GATGCTACCT  | 22320 |
| AGGCAAGTCC | AGATATATAA  | TCCGGGCGTG | AGAAACAGAA  | ACGGCCAATA  | ACCCATGTTT  | 22380 |
| TTGGAAAAAC | ATCAGACACC  | TTAACACAAA | TCATGTACAC  | CTGGTATTAC  | TATTTCCCAAC | 22440 |
| ACATCTTATA | GCATTTCAAA  | GATAAGGGTG | CCTTACGGGG  | CGCCCGAAAC  | AAGTGGGCGG  | 22500 |
| GCCTACTTCA | CTGTTTATAA  | GTCAGCCGGA | CCAAGCTGCT  | GCTCTTGGGG  | ACGTGACTGC  | 22560 |
| TTCTGGGCGC | AGCTGCCTCC  | AAATGATACA | CACATTTTTT  | GATTGTCCCG  | GGCGCCGGCT  | 22620 |

|            |             |             |             |             |             |       |
|------------|-------------|-------------|-------------|-------------|-------------|-------|
| AGTGGAGGGC | GGAGTTATAT  | CAAGCTACTT  | TCTGATTGGT  | GCCCCAGGCA  | GGACTGCCAT  | 22680 |
| AAAAACTGAA | GAAGGCGTGT  | CTGCTTTGCA  | GAATTTACCC  | CCCACTGTGC  | TCCCGGTTGC  | 22740 |
| TGGCACCGBT | TCAGTGGTCC  | GACCTGTCTT  | CTGTGCTCCC  | CCGTGGACGA  | CGCCGASTGC  | 22800 |
| CTCTCGGGGG | TCCATGTCTA  | GCCTCTTCAT  | TTCAATTACCT | TGGGTGGCGT  | TCATCTGGCT  | 22860 |
| AGCCCTCCTT | GGCGCGGTTG  | GGGGTGCCCG  | CGTTCAGGGG  | CCCATGCGGG  | GCTCTGCTGC  | 22920 |
| CCTCACCTGC | GCCATCACGC  | CCCCTGCTGA  | CATAGTTAGC  | GTTACCTGGC  | AAAAAAGGCA  | 22980 |
| GCTCCCCGGT | CCCGTAAACG  | TGCGCACGTA  | CAGCCATTCA  | TATGGGGTGG  | TGGTTCAGAC  | 23040 |
| CCAGTACCGC | CACAAGGCAA  | ATATACCTG   | TCCTGGGCTT  | TGGAACTCTA  | CCCTTSTTAT  | 23100 |
| CCATAACCTT | GCASTGGATG  | ATGAGGGCTG  | TTACCTGTGT  | ATCTTTAACT  | CATTTGGTGG  | 23160 |
| CCGGCAGGTG | TCATGCACAG  | CCTGCCTGGA  | AGTGACATCT  | CCCCCTACTG  | GACACSTGCA  | 23220 |
| GGTAAATAGC | ACAGAAGACG  | CAGACACCGT  | CACCTGTTTG  | GCAACTGGTC  | GCCCCCCCCC  | 23280 |
| CAATGTCACC | TGGSCCGCAC  | CCTGGAACAA  | CGCCTCTTCT  | ACCCAGGAGC  | AGTTCACTGA  | 23340 |
| CAGTGATGGT | CTTACAGTTG  | CGTGGAGGAC  | CGTGAGGCTG  | CCGCGTGGGG  | ATAATACCAC  | 23400 |
| CCCAAGTGAG | GGAATATGTC  | TCATCACCTG  | GGGAATGAG   | AGCATATCAA  | TCCCGGTTTC  | 23460 |
| TATTCAAGGC | CCCTTGGCCC  | ATGACCTTCC  | CGCGGCCCCAG | GGAACTCTTG  | CCGGGGTTGC  | 23520 |
| CATTACTCTG | GTGGGCTTAT  | TTGGGATATT  | CGCATTACAT  | CATTGCCGCG  | GCAAGCAGGG  | 23580 |
| CGGTGCATCA | CCTACTTCAG  | ATGACATGGA  | CCCCCTATCC  | ACCCASTGAC  | TAGATGGACA  | 23640 |
| CCCCGTGAAC | CSTCGTGCTT  | ACCCACCCCC  | TTCTGATTCT  | GACAGACAAC  | ACTACTATGT  | 23700 |
| CCCAAAGACT | GTTTTTTTACA | GCCCCGATGGC | CCTTCAGGCC  | TCCTTGAGTG  | TCTAGCTGGT  | 23760 |
| CCCGTGSTCA | TTGTGTGGTT  | TGGCAGTCAC  | TTCCCCATTT  | TGGTGTGGGG  | TTTTGGGTTT  | 23820 |
| TGCCCTGCCC | CCAGGCCAACG | TGGATCATAT  | TCTTTCCCGT  | CAGGGGAGTG  | ACAAAGCTATA | 23880 |
| GGACAGAAAG | GTCACCTGGC  | CCAAACGGAG  | GATCCTAGGT  | GGGTGTGCAT  | TTATTAGACG  | 23940 |
| TTGGTGTGTT | GAAGGACGGA  | TCAGGCGGGG  | AGGAGGGGGT  | GGGGGAGAGT  | TACTGCAGCA  | 24000 |
| CTAGGTTAGG | TTGAAAGCCG  | GGGTAAAAGG  | CGTGGCTAAA  | CAACACCTAT  | ACTACTTGTT  | 24060 |
| ATTGTAGGCC | ATGGCGGCCG  | AGGATTTCCCT | AACCATCTTC  | TTAGATGATG  | ATGAATCCTG  | 24120 |
| GAATGAAACT | CTAAATATGA  | GCGGATATGA  | CTACTCTGGA  | AACTTCAGCC  | TAGAAGTGAG  | 24180 |
| CSTGTGTGAG | ATGACCACCG  | TGGTGCTTAA  | CACGTGGAAC  | GTTGGAAATAC | TCTCTCTGAT  | 24240 |
| TTTCCTCATA | AATGTTCTTG  | GAAATGGATT  | GSTACCTTAC  | ATTTTTTTGCA | AGCACCGATC  | 24300 |
| GCGGGCAGGA | GCGATAGATA  | TACTGCTGCT  | GGGTATCTGC  | CTAAACTCGC  | TGTGTCTTAG  | 24360 |
| CATATCTCTA | TTGGCAGAAG  | TGTTGATGTT  | TTTTTTTTCCC | AATATCATCT  | CCACAGGCTT  | 24420 |
| GTGCAGACTT | GAAATTTTTT  | TTTACTATTT  | ATATGCTTAC  | TTGGATATCT  | TCAGTGTGTT  | 24480 |
| GTGCTCAGT  | CTAGTGAGGT  | ACCTCCTGGT  | GGCATATTCT  | ACGCGTTTCT  | GGCCCCAGAA  | 24540 |
| GCASTCCCTC | GGATGGGTAC  | TGACATCCGC  | TGCACTGTTA  | ATTGCATTGG  | TGCTGTGGGG  | 24600 |
| GGATGCCTGT | CGACACAGGA  | GCAGGGTGGT  | CGACCCGGTC  | AGCAAGCAGG  | CCATGTGTTA  | 24660 |

|            |            |             |            |            |             |       |
|------------|------------|-------------|------------|------------|-------------|-------|
| TGAGAACGCG | GGAAACATGA | CTGCAGACTG  | GCGACTGCAT | GTCAGAACCG | TGTCAGTTAC  | 24720 |
| TGCAGGTTTC | CTGTTACCCC | TGGCCCTCCT  | TATTCTGTTT | TATGCTCTCA | CCTGGTGTGT  | 24780 |
| GGTGAGGAGG | ACAAAGCTGC | AAGCCAGGCG  | GAAGGTAAGG | GGGCTGATTG | TTGCTGTGGT  | 24840 |
| GCTGCTGTTT | TTTGTGTTTT | GCTTCCCTTA  | CCACGTACTA | AATCTACTGG | ACACTGTGTT  | 24900 |
| AAGGCGACCG | TGGATCCGGG | ACAGCTGCTA  | TACGCGGGGG | TTGATAAACG | TGGGTCTGGG  | 24960 |
| AGTAACCTCG | TTACTGCAGG | CACTGTACAG  | CGCCGTGSET | CCCCTGATAT | ACTCCTGCCT  | 25020 |
| GGGATCCCTC | TTTAGGCAGA | GGATGTACGG  | TCTCTTCCAA | AGCCTCAGGC | AGTCTTTTCT  | 25080 |
| GTCCGGGCGC | ACCACGTAGC | CCGCGGATGT  | CTACGTGCCC | TTCCCCCTTA | ATTTAATCTA  | 25140 |
| GCCTCCGSET | CCCATGATGC | AGAGAGGCGA  | ATTTGGTTTT | TACACAGATG | TGACTATGTA  | 25200 |
| TTTGTGTTAT | TATGCGATTA | AATGAGGGGT  | CTGATCCCAA | AAGCAATGTT | TAGTGGTGGT  | 25260 |
| CGTTGATCTT | CTTGACGCTC | CATAGGTAGA  | TTGACTGGAA | CGCCATGGCC | CACGGGGACA  | 25320 |
| TGGACAGGGG | TGTTAGGTCT | GGTGGAAACAT | GCTGCCACTG | CCACGGATGG | AACATCAGAG  | 25380 |
| ATGGGTCTAT | GATCAGGGCA | GCGTGTCCGC  | CGTCACTGGA | TGTAAGTCCG | GCCACCGTGG  | 25440 |
| AGTTGCCTGT | GGGGTTTCTG | GGATAGTGTC  | TGGCTGGCAG | GGTCTCATCC | GCGGCATTTG  | 25500 |
| CATGCTAGGT | GAGGGTTATC | TGCGCTCGCT  | GTCTCAGTAT | GTACTCGAGG | GCGTCCCTGT  | 25560 |
| CGTACCGGAC | CCCCAGGTAC | TCTCCCTGGG  | CCCAGCTGGG | CAGCACCGTC | CCCCGCAACA  | 25620 |
| CTCGGAGGAA | AACGCTCTTA | GTGTTCTGAG  | GGATCTGTAT | GTTTAGCCAG | TGGCTGTCTAT | 25680 |
| ACAGCTTGGA | CACGTTGGTC | TCCAGGTTTA  | CCGCCCAGCG | CTGGGCTGGT | GTGGGTCCGT  | 25740 |
| ACGTGTATGG | TGAGGATTCC | GACCGGCCCA  | CTACACCCAG | GGCCACCAGC | AGCTGGAAGC  | 25800 |
| CCACCTCGCC | ACAGCAGATG | GAGAAATGTGT | CGGGTCTGTT | TAGAAACTCT | GTGAGGGTGG  | 25860 |
| AGGCACAGGT | AGGGTCGTTA | CACAGCGCCA  | GGACCCATCC | CCTGGCGCTG | GCGTAGCTGG  | 25920 |
| CCTGGCAGCC | TGTTCTGAGA | CATGTAATCA  | GACCAGAGAA | CCCCGACAAG | GACTGTCTCT  | 25980 |
| GTTTAAGCTC | TTCCACAGTC | ACCGTGCGCA  | CCTCAAGGCC | CGTGTCTCTG | AACGTTGGCA  | 26040 |
| TGAGCGCGTA | CGGGGCACTG | CTCCCAAGCA  | GCACCAACCG | GGCCACACCG | CGTGGGGAGG  | 26100 |
| TGGGGCACGA | AAACAGGCGC | AGCTGACTCC  | CAAGGCACAT | GGCCCTTAGG | CTGCCCAGGT  | 26160 |
| GATGCTCCAG | ACGACCCAGG | TCCTTCCTGT  | GCATGTCTCT | CAGTGGGTGC | AGGGGAGGCG  | 26220 |
| TCACCAGGTT | CCACATTTCC | TCAGAAAAGG  | AGGTCCATGA | GACTTGCAAG | GAACTCAGGG  | 26280 |
| TCTCTTGAAA | CACAACTGTC | TGTTCTTGCA  | AAACCGTGAC | GTTGTTGCCT | TCTCCCTCGG  | 26340 |
| GGCCAACGGT | GCCCAAGTGG | TGTGCCACGC  | AGCGGTAGTC | CCTGGCCGCG | CGCAGCACCT  | 26400 |
| CTGACAAAGT | TACCTGGGGC | ACCTCAACCA  | GTGCCCCAGG | GGTCTCTGAA | ACCATAAAGT  | 26460 |
| CGAGCGGGTT | AGGGTGGGCG | GGTAGTGAGA  | GCTGCAGTCC | CCTGCAGCCG | GCCAGGGCCA  | 26520 |
| TCTCGATTGC | AGATGGGAGA | AGCCCTCCGT  | CCCCTATGTC | GTGCCCAGAT | ACCAATGAGCC | 26580 |
| TCTTGGACAT | CAGGTACTTA | ACAAGCATGA  | ACAGGCTGGC | GACCCTGGAC | GGGTTTCAGG  | 26640 |
| GGGGTATTGG | GTGCCTGGAT | GCCAGGAAGT  | TGTGCTCGAA | GSTGGACCCG | GCTATGAGAT  | 26700 |

|            |            |            |            |             |            |       |
|------------|------------|------------|------------|-------------|------------|-------|
| AGCTCTGATT | CACGGCCAGG | TATACCAGGG | CGTTGCCTTC | GACCTTTACG  | TCCGGGGTGA | 26760 |
| CCCTGTATCT | GGATCCCTTG | ACCTCGGCCC | AGCTGGTAAA | CACCACCGAG  | TTGAAGGGAA | 26820 |
| GGACCTCCAC | CGTTTCTTGC | TGTTGTGTGA | TGCGCACATG | GCGCTCCGAA  | AGCGTCCGAG | 26880 |
| AGCTGGCAGC | CGAGGAGATG | GACAGTGCCA | CTCCCAGCTC | CCGGCAGAAAT | TCCCTTGACG | 26940 |
| CGAAGAGGCA | CTCCTGTAGG | AGGCCGGGCT | GGTGGTCCTC | TGGACTCCAC  | GCCACGGGCG | 27000 |
| CAGTTAGCAC | TACGTCTTGG | AGCTTGACAA | CGGGACTGAA | CATGAGGTTG  | GTGAGAGGCT | 27060 |
| CGGTGATGGC | ATAGGTGGCC | CCGGTGGATA | CATTAGTAGC | CATCTTGTAG  | GCCTGCTCCG | 27120 |
| CCATGGCCAT | TGCCTGACCC | CTCCACGCTG | GCACTGGAAG | CAGCTCCTGG  | GGCAGGGGCT | 27180 |
| TCACCCAGGT | CTCGAAGTCC | TTGTGTAGGA | GGTTGGCCAT | GGACGGAGTG  | ATGGCCTCCA | 27240 |
| CCGTGTCCGG | CACCTCTGGC | GCCACCCCTC | CGGCCAGCAT | GGACGAGTGC  | AGCACCAGGT | 27300 |
| GGTAGTCTGA | AACCGGTATG | TCCAGGGGTC | CCACGCCAGC | CTGTTGGGCG  | ATGAGGGGCT | 27360 |
| TGGAGCATCG | GTCCATGTGT | CGCGTAAAGA | ACTCCTTGCT | GCCAACCGTC  | GAGTGGCGAA | 27420 |
| GTAACCTGGT | GATTGTGGAG | CCGGTGGCAA | AAAGGCCCCA | GTCAACATCC  | TCCGGGTGCC | 27480 |
| CCGAGACCGG | GACACCATCG | GACAGCGCCA | GCCAGGGGGA | CGGGGGGGTG  | GACGACGGCT | 27540 |
| GGTCTACAGA | GAAGACCCCT | GTGGTCTCCC | CGGTACAGTC | GTCTACTATT  | CTGATGCTTG | 27600 |
| GGTGCTCCGA | GGTCTTCCCG | AGGACCGTTA | CCTGGCACGC | GCACAGGCGC  | GCGGGGCGCT | 27660 |
| GCAGTACCTC | CAACGGGGTC | TCCGCCAGAT | CCCCAGGCAC | CGCGCCCGAC  | TCTGCCACCA | 27720 |
| CCGCAAAACG | CAGGGAGCAA | TACACGTTGA | GAAAGTGCTC | TGCCACCGCC  | GCCTTCACGG | 27780 |
| CATCCGGACC | GGCCCGGGGA | TCCGCAGGCA | GGTGGGTGCC | CACCTCGTCC  | GCTAGCTTGG | 27840 |
| AGACAAACAG | CTCCAGGGCG | GTCCGCGGGC | CCAGCGCTTG | CAGGTGCCTC  | ACCACGGGGG | 27900 |
| CCGGGTGATG | CGATCTGTTT | AGTCCGGAGA | AGATAGGGCC | CTTGGCAAGC  | CGGTGGACCA | 27960 |
| GCTTCAGGGT | CTCCAGAGTG | CGCACCGCAT | TGTCCGAGCT | GTCCCGATAG  | AGGTAGGGGT | 28020 |
| AGGTGTCCGG | TCCATCCGTG | GGCTCAAAAC | TGCCCAGACA | CACCACTGTC  | TGCTGGGGGA | 28080 |
| TCATCCTTCT | CAGGGAGATG | CATTCTTTGG | AAGTAGTGCT | AGAGATGGAG  | CAGACTGCCA | 28140 |
| GGGCGTTGCC | AGGAGTGGTG | GCGATGGTGC | GCACCGTTTT | TAAGAAACCC  | CCCAGGGTGG | 28200 |
| GGACTCCCGC | TCCCTTCAAG | ATCTCGGCTT | GCTGTACGCC | CTTGGCGAAT  | ATGTGACGGA | 28260 |
| ATCGGCTGTG | CGCACGGGGT | CCCAGGGGCG | GTTGGGTGGT | ATACAGGGCG  | GTGAGGGGCG | 28320 |
| CCTGTGTCTG | TCCGCTTGGA | AACAGGGTGC | TGTGAAACAG | CAGGTTGCCA  | AGGCGCGGAA | 28380 |
| TACCCCTCTG | CACGCTGCTG | TGGACGTGGG | TGTACGCTCC | GTGGATCCCG  | AACGGCTGTC | 28440 |
| TGGCACAGTT | CCAGGGCCAC | CGTTCCATGG | TGCATCTTCC | CGGTATCACA  | AAGTACCTGG | 28500 |
| CCACGTTATA | ATTGTCCCGG | GTTGAAGGCT | GCACCGCCAG | CGGTAGCAGG  | TTTGTCCCGA | 28560 |
| GGGATATCAT | AACAGGCTGC | ATAATGACAT | CATCTTCAAT | GTGTGGGCTA  | GCCACGGGCT | 28620 |
| GGGGACCCCT | GGGCACTTCC | AACCCCTCTT | ACGGTACCAG | GTGGGTATTT  | TGTGTAAATG | 28680 |
| CCCTGATAAA | CTGAGGTGGG | TGTGGTTCTA | GCAGGGTCTG | TGTGATTTTG  | GACACCAAGT | 28740 |

|             |            |             |            |             |             |       |
|-------------|------------|-------------|------------|-------------|-------------|-------|
| GCCTGCCAC   | TTCCACTCTA | GCCCACTCCT  | GCAATCCTAG | CTCTTGCCAGC | AGAACTGCCAA | 28800 |
| GCTTGTGGA   | CAATGTTGTG | GGCCGGTGGT  | GCATGTTTGG | CCCGTAGCCA  | AAGGATACAA  | 28860 |
| CACGTCGCT   | CCCCCGTGGC | ACAGACCGCC  | TGATGACATG | GGGATATCCA  | AGGAGCGGTG  | 28920 |
| ACAGCACAGC  | GAGCACCGTC | TGTATTTCCA  | CATCCCGTCT | CTCTCGCTCC  | TCCCTCGAAG  | 28980 |
| TGGGAGGTCT  | TCGGAAAGTT | ATCCATAGCA  | GATAGTAGCC | TCCGGTGCCA  | CCGGGTACGA  | 29040 |
| GAGTGAGTGT  | GCCCGTACGG | CTTGTATAAA  | AGTTCACAAA | AGCTTCCTCA  | TCCGCGGTGA  | 29100 |
| GATCACTCTC  | CAACCACAGC | CCAGTGACGT  | CGTAGGCCAT | GCCTAGAGGG  | CGCACCGCCC  | 29160 |
| CCGGGGACAC  | CCTCTGTAGT | CAGGCTGCCG  | AGAAACCCGC | GAGATCTCTG  | GGGAGTAGGA  | 29220 |
| AGAAACTTAG  | AATCCCCAAA | TATGTCGCAG  | TCACAGGTTG | TCGGGCAGAG  | TCTGTTTCCG  | 29280 |
| CTTTCATGGG  | ATCCACAGTT | ACTTGTAGCC  | ATGTCACTAA | CCTCAAATAC  | TCAAAAAAAG  | 29340 |
| CTATCGATGG  | AAAAATGCTG | TGGTCCTAGG  | TTAGTCCGTG | GGAAACAAAA  | CTTCCTCATA  | 29400 |
| CACCTTCATCT | GCAGGCTGAA | ATGGTGGCGG  | ATCCAGACTC | CTTACACCAC  | AGTTGCTCAC  | 29460 |
| ATTAGAGATA  | CCTGATTGGT | TAATACAAGC  | GGACGCACGC | GTTGCTGGAG  | GCGTGTGTCT  | 29520 |
| GCCCCAAGATA | CTAGCATAGG | TGACTGTGCG  | TTGCTATGT  | AGTTGCTGCA  | TTTCAAGTTG  | 29580 |
| GGTCGTTACT  | TCTGTGTTGC | AAACCCCTTAC | TGGAGATAAT | GCCATGTCTG  | TTGTGGAACT  | 29640 |
| TAAAATACGC  | GAGTGATATA | CATTTCTAGA  | TGGTAGAGGT | GGTAAACGGC  | GAGCTAAATG  | 29700 |
| ATTAACATCG  | GGACATATCC | TGCCTGCATG  | AGCATGTGGT | GTGTCTGTGT  | GTGTATATAT  | 29760 |
| TGGTAATCTT  | GTTGTTACAT | TGTTGAACGA  | CACAAGTCTG | CTCTCTCGGT  | AGAGATAACC  | 29820 |
| CACCACTAGG  | GCTTGGCCAG | TACCTAATAA  | GAAAAAATAA | AATCGTTAAT  | CTCTTTTTTT  | 29880 |
| ATGTGGCGCT  | GSTGTTCCAA | TTATAAATAA  | AAACACAAC  | CACCTAATAT  | CACAATTACA  | 29940 |
| CAATCAGTC   | CTGAAGTAAC | ACCTGTAGTC  | CAACCGTCAG | TGTAGAGCAG  | GACTAACTTA  | 30000 |
| ACACAGCATC  | CAGCACATGT | CCATGCTAAG  | GAAATAAACC | AAAGTTATGT  | TTGGSTTTGC  | 30060 |
| TTTATGACCA  | GGGAGCTGCT | ACCCAGGTAC  | AAAAAATCCT | TACCCAAAAA  | TAGAAACAGG  | 30120 |
| AAGCCACCAG  | AGAGTGAAGC | TTTGTGAAG   | CTTTGCCAGC | AGAAGAAACA  | ATATAATAAA  | 30180 |
| AAGCCACAGC  | CTGCTAGTAA | TGTTATACTC  | CCTGTAAATA | AAAAATATGG  | ACAGTAATAA  | 30240 |
| TTTATGACAC  | CCAATAAGTA | TGTGGAAAAA  | ATGTAATGTA | AACCACTATA  | CTGTAATAAA  | 30300 |
| CATACCTTCG  | TTATTGGTGT | CTTGTTCGGG  | CTTTATAAAC | AGTATCCCTA  | TTGTTGTGGT  | 30360 |
| TAGTGTAACC  | AACACTCCTC | CTTGTAAGAG  | TAAAAATGAC | ATAAGCCCCCT | TAGTTGATCC  | 30420 |
| AATCCAATGT  | CGTTTCATTG | TTATAAACAA  | GCCGGTCATA | CCTGTAATAA  | AGTTATTTCAT | 30480 |
| TACAAAATGT  | TATAATAGTA | TTGGTAATGT  | TTAGTTAAGA | TAATGTAAAC  | TTACAGTAG   | 30540 |
| TCATATACCA  | ATATGTATGC | AGCTTATGCA  | TCCTGCGATG | ATTACASAAA  | GGCATGAATG  | 30600 |
| GGAAACGCCA  | AAAAAGGCCG | GTGTTGCCCT  | GAGTATACCT | GTAATAAAAA  | ATAAATAATA  | 30660 |
| TTGTTGGTTG  | CAATGCTTAG | GTGCAAGCAG  | ACATAATTGC | ATAGCAGTAA  | AAACCAGACT  | 30720 |
| TACCACCACA  | TATTGCAAAC | ACACATGCAG  | CGAGCTTGAG | ACAAGGCCCA  | TTATCTGTTG  | 30780 |

|             |            |            |            |            |            |       |
|-------------|------------|------------|------------|------------|------------|-------|
| CAAAGATATG  | TATAAAAAA  | ACAAGCAACA | ATGTCCATAA | TGGCAAAAAA | AAGTGGCAAT | 30840 |
| GTGTCCAGTT  | GTTGTAAATC | TGCAATCCCA | TTGAGAATAT | AAGTACCAAC | ACCATAACAA | 30900 |
| TGCACAGTAA  | TCCGCTATCA | ATAGTGCAAT | TAACGACTCT | TAATGTTCCA | CCAASTGATA | 30960 |
| GAATGGCTGA  | AAAACACATA | CAGGGGAATT | ACGTTTTTTT | AAAAAATTGG | AAATATTAGA | 31020 |
| TACATAATTT  | TTATTTAATA | AAAAACCTTT | AGTAAACTTT | ACCAGTAATT | ATAGACAATA | 31080 |
| AAGTTATAAT  | ACAAACACAA | ACAGTACTCA | AAGTACTTTG | AGTAGAGAAA | CTCCAACTGG | 31140 |
| CAAAGGCAAT  | ACATCCTAAA | ACAAAAGACA | AATACACGAG | ACATTTAAAC | AATGTATACT | 31200 |
| TAGAAAGAAA  | TAAGTTAAAC | ATTTAAAAAA | TGTAACCTAC | CAACAATTAT | AGATGSTCCA | 31260 |
| ATGGGAGGGG  | AAGCTTGAAA | ACGTTGTTTT | TTTGACTGCA | CATATATGTT | GTTATTGTAC | 31320 |
| AAAAAGTTG   | GTAGTAAACA | CTTATGTTAC | TGAGCAAAAA | TATGGTGTTT | TGTAAATTTA | 31380 |
| TAGTTAAAAG  | ACAAAACATA | ATAGACAAAC | ACCCACAACA | TGTTATAAGT | GCTGCAAACC | 31440 |
| AAGTACCCCA  | CAGGTATTTT | TTGTAATTCA | TTGTAGACAA | AAAGCCCCAG | GCCCCAAAAT | 31500 |
| GAAGTGGACA  | AAAGAAATAT | GTAATTAAGT | GTAGTTGGAC | AAGGAATTAT | ATAGCTGGAT | 31560 |
| GAGTTAGTTT  | TGCACAGAAC | CAGACATCCT | ATTTTTGTTT | GGAAACCTAA | AATCCGGATG | 31620 |
| AAGGGCTTAT  | AAAATGGCAC | AGCTGCAAAA | AGCTGATAAT | GTAACACTGC | ATCCTGCTGT | 31680 |
| TTTTGATTGT  | AGCGGAAAAA | TGTAATAAAT | TTTACAGACA | GTTTTGCCTA | CTGAGAACAT | 31740 |
| GTTGAAAAAA  | AGGCACTAAG | GGCTTTTTTG | CCAAAGGAAA | AATGCCCCCG | TGGGGTTAGG | 31800 |
| GGAAAGGGGG  | GATGGGGTGA | TGGGGGAATG | GTGGGAAAGG | GGGATGGGGG | TGATGGGGGA | 31860 |
| ATGCTGGGAA  | AGGGGTGATG | GGGTGATGGG | GGAATGGGGG | GAAAGGGGGA | ATGGGGGGAA | 31920 |
| AGGGGGGAATG | GGGGGAAGG  | GGGAATGGGG | GGAAAGGGGG | GATGGGGGGA | AAGGGGGGAT | 31980 |
| GGGGGGAAAG  | GGGGAATGGG | GGGAAAGGGG | GGATGGGGGG | AAAGGGGGAA | TGGGGGGAAA | 32040 |
| GGGGGGATGG  | GGGGAAACGG | GGGATGGGGG | GAAAGGGGGG | ATGGGGGGGA | AAGGGGGGAT | 32100 |
| GGGGGGGAAA  | GGGGGGATGG | GGGGGAAAGG | GGGGATGGGG | GGGAAAGGGG | GGATGGGGAA | 32160 |
| GGGGGGGGGG  | AGGGGGAAAG | GGGTGAAGGG | GGAAGGGGGG | AGGCGAA    |            | 32220 |



What is claimed is:

1. An isolated nucleic acid encoding a Kaposi's sarcoma-associated herpesvirus (KSHV) polypeptide.  
5
2. The isolated nucleic acid of claim 1 wherein the polypeptide is selected from the group consisting of:  
10
  - a. Thymidylate synthase (TS);
  - b. Viral protein kinase;
  - c. Alkaline exonuclease (AE);
  - d. Helicase-primase, subunit 3; and
  - e. Uracil DNA glycosylase (UDG).
- 15 3. The isolated genomic DNA molecule of claim 1.
4. The isolated cDNA molecule of claim 1.
- 20 5. The isolated RNA molecule of claim 1.
6. The isolated nucleic acid molecule of claim 1 which is labelled with a detectable marker.
- 25 7. The isolated nucleic acid molecule of claim 6, wherein the marker is a radioactive, a colorimetric, a luminescent, or a fluorescent label.
- 30 8. A replicable vector containing the isolated nucleic acid molecule of claim 1.
9. A host cell containing the vector of claim 8.
- 35 10. The cell of claim 9 which is a eukaryotic cell.
11. The cell of claim 9 which is a bacterial cell.

12. A plasmid, cosmid,  $\lambda$  phage or YAC containing the isolated nucleic acid molecule of claim 1.
- 5 13. A nucleic acid molecule of at least 14 nucleotides capable of specifically hybridizing with the isolated nucleic acid molecule of claim 1.
- 10 14. An isolated thymidylate synthase polypeptide the sequence of which is set forth in SEQ ID NO:1.
- 15 15. The isolated polypeptide of claim 14, wherein the polypeptide is linked to a second polypeptide to form a fusion protein.
16. The fusion protein of claim 15, wherein the second polypeptide is beta-galactosidase.
- 20 17. An antibody which specifically binds to the polypeptide of claim 14.
- 25 18. The antibody of claim 17, wherein the antibody is polyclonal antibody.
19. The antibody of claim 17, wherein the antibody is a monoclonal antibody.
- 30 20. A host cell which expresses the polypeptide of claim 14.
21. A vaccine which comprises an effective immunizing amount of the polypeptide of claim 14 and a suitable pharmaceutical carrier.
- 35 22. An antisense molecule capable of hybridizing to the isolated nucleic acid molecule of claim 1.

23. The antisense molecule of claim 22, wherein the molecule is a nucleic acid derivative.
24. A triplex oligonucleotide capable of hybridizing with a double-stranded isolated nucleic acid molecule of claim 1.
25. A transgenic nonhuman mammal which comprises the isolated nucleic acid molecule of claim 1 introduced into the mammal at an embryonic stage.
26. A method of diagnosing Kaposi's sarcoma comprising: (a) obtaining a nucleic acid molecule from a tumor lesion or a suitable bodily fluid of a subject; (b) contacting the nucleic acid molecule with the labelled nucleic acid molecule of claim 6 under hybridizing conditions; and (c) determining the presence of the nucleic acid molecule hybridized, the presence of which is indicative of Kaposi's sarcoma in the subject, thereby diagnosing Kaposi's sarcoma.
27. The method of claim 26 wherein the nucleic acid molecule from the tumor lesion is amplified before step (b).
28. A method of diagnosing a DNA virus associated with Kaposi's sarcoma comprising: (a) obtaining a suitable bodily fluid sample from a subject; (b) contacting the suitable bodily fluid of the subject to a support having already bound thereto a Kaposi's sarcoma antibody of claim 17, so as to bind Kaposi's sarcoma antibody to a specific Kaposi's sarcoma antigen; (c) removing unbound bodily fluid from the support; and (d) determining the level of Kaposi's sarcoma

antibody bound by the Kaposi's sarcoma antigen, thereby diagnosing Kaposi's sarcoma.

29. A method of diagnosing a DNA virus associated  
5 with Kaposi's sarcoma comprising: (a) obtaining  
a suitable bodily fluid sample from a subject;  
(b) contacting the suitable bodily fluid of the  
subject to a support having already bound thereto  
10 a Kaposi's sarcoma antigen encoded by the  
isolated nucleic acid molecule of claim 1, so as  
to bind Kaposi's sarcoma antigen to a specific  
Kaposi's sarcoma antibody; (c) removing unbound  
bodily fluid from the support; and (d)  
15 determining the level of the Kaposi's sarcoma  
antigen bound by the Kaposi's sarcoma antibody,  
thereby diagnosing Kaposi's sarcoma.

30. A method of treating a subject with Kaposi's  
sarcoma comprising administering to the subject  
20 an effective amount of an antisense molecule of  
claim 22 under conditions such that the antisense  
molecule selectively enters a tumor cell of the  
subject, so as to treat the subject.

31. A method of treating a subject with Kaposi's  
sarcoma comprising administering to the subject  
25 having a human herpesvirus-associated KS a  
pharmaceutically effective amount of an antiviral  
agent in a pharmaceutically acceptable carrier,  
wherein the agent specifically binds to the  
30 polypeptide of claim 14.

32. A method of prophylaxis or treatment for Kaposi's  
sarcoma (KS) comprising administering to a  
35 subject at risk for KS, the antibody of claim 17  
in a pharmaceutically acceptable carrier.

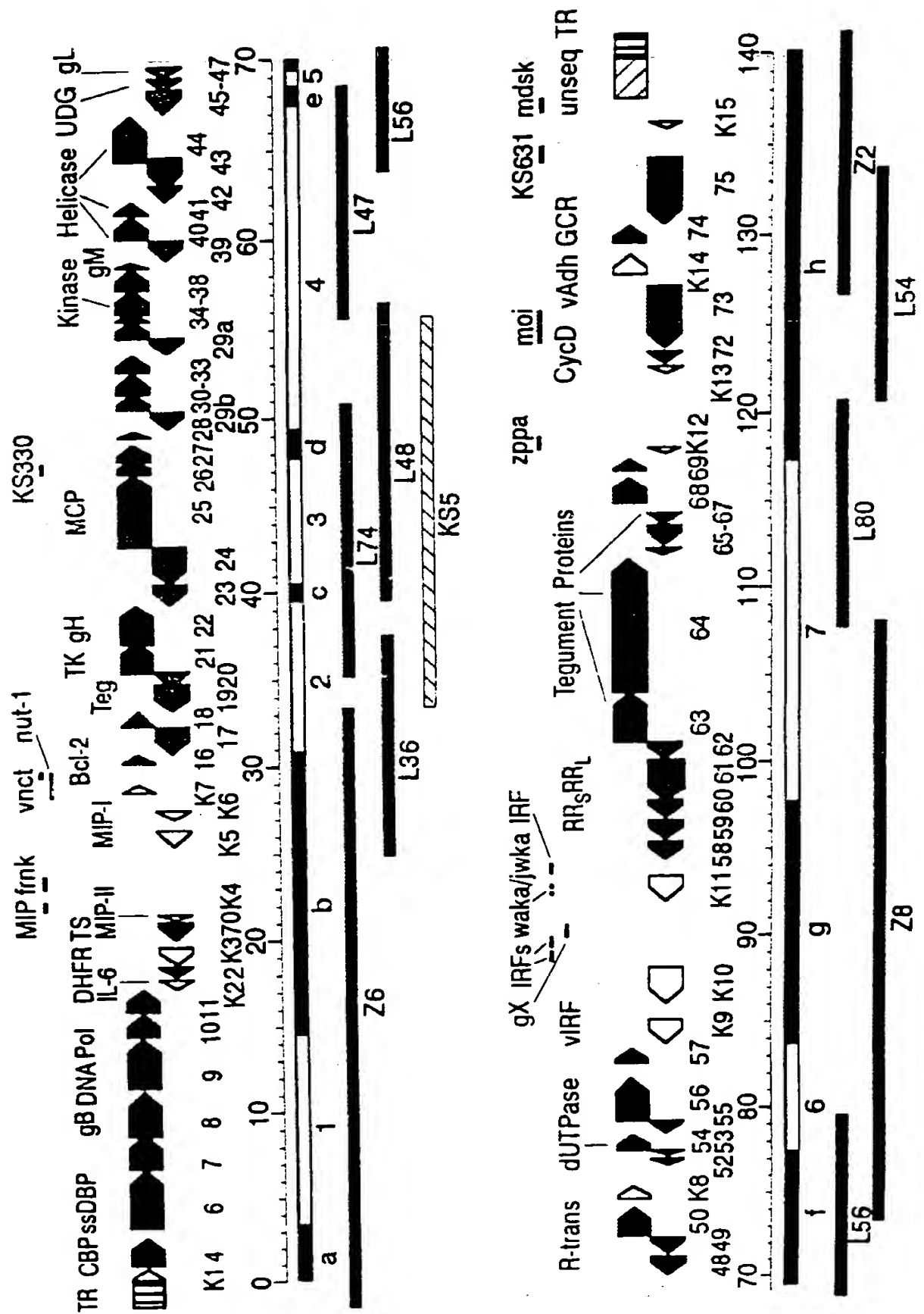
33. A method of vaccinating a subject against Kaposi's sarcoma comprising administering to the subject an effective amount of the polypeptide of claim 17, and a suitable acceptable carrier, thereby vaccinating the subject.
34. A method of immunizing a subject against a disease caused by the herpesvirus associated with Kaposi's sarcoma which comprises administering to the subject an effective immunizing dose of the vaccine of claim 21.
35. A method of identifying a compound that inhibits KSHV replication in a subject which comprises:
- (a) expressing a KSHV enzyme in a bacterial auxotroph, which auxotroph is dependent on a product of the expressed KSHV enzyme for bacterial growth;
  - (b) administering the compound to the auxotroph;
  - (c) measuring bacterial growth; and
  - (d) comparing bacterial growth in step (c) with that of the auxotroph in the absence of the compound so as to identify a compound that inhibits KSHV replication in the subject.

36. The method of claim 35, wherein the KSHV enzyme comprises one from the list as set forth in claim 1.

37. The method of claim 35, wherein the KSHV enzyme comprises thymidylate synthase.

1/15

FIG. 1



2/15

FIG. 2A

1 CGTGAACACC CCGCGCCCCG CGCCCCCCAC ACCGCGCCGC CCTCCCCCT CCCCCCGCTC  
 61 GCTCCCGC GCTGCCGCCA GGCCCGGCC GGAGCCGGCC GCCCGCGGG GGCAGGGCGC  
 121 GCCCGCGGC TCCCTCGCGG GCGGGGAC GGGGAGGg ggcgcgggc CCCC CGCGC  
 181 CGGCGAGCG GAGCGGAGc gccccgcgc CTAGgcccag cctactttt tttcgggcg  
 241 CCGAGCCCCG AGCCCCGCCG GGTACGGG CCGCGGCC CCGCGCGCC CCGCGGGG  
 301 gccccgcac cctctcgc cccccggTC CCGCGGCC CCGCGCGCC CCGCGGGG  
 361 GTAAACAGG GGGGGGGA TCGGCGCG GCGCGGCC CCGCGCGGC GCGCTTGCT  
 421 ttcgtttct cccgcggccc cccgggcgc agccgcgcg cggcggcgg cgcctcc  
 481 cccgggggc tcggcgggg gccCCTGTC CCcgcgcgg cccgcgacc cCGCGCCG  
 541 CGCGCCCCGA TCCCGCGGC GCCCGCCCC CCTGCCGGG ACGCGCCGG CCTGCGGG  
 601 CCTCCGCC GGCATGGG cgcgcgcgc ctcaggcc cggcgcggc ggcgcctgt  
 661 cccgcgcgc gccgcgggg gaccccgGC AGCGAGGGA GGGCGGCC TCTCTCTACT  
 721 GTGCGAGGAG TCTGGGCTGC TGtGTGTGAG CCTGTtTGGG GGAGCCTCCT CAGTGCTTGC  
 781 TACGTGGAGC CCTGGACACT A

3/15

FIG. 2B

Nde II

1 2 3

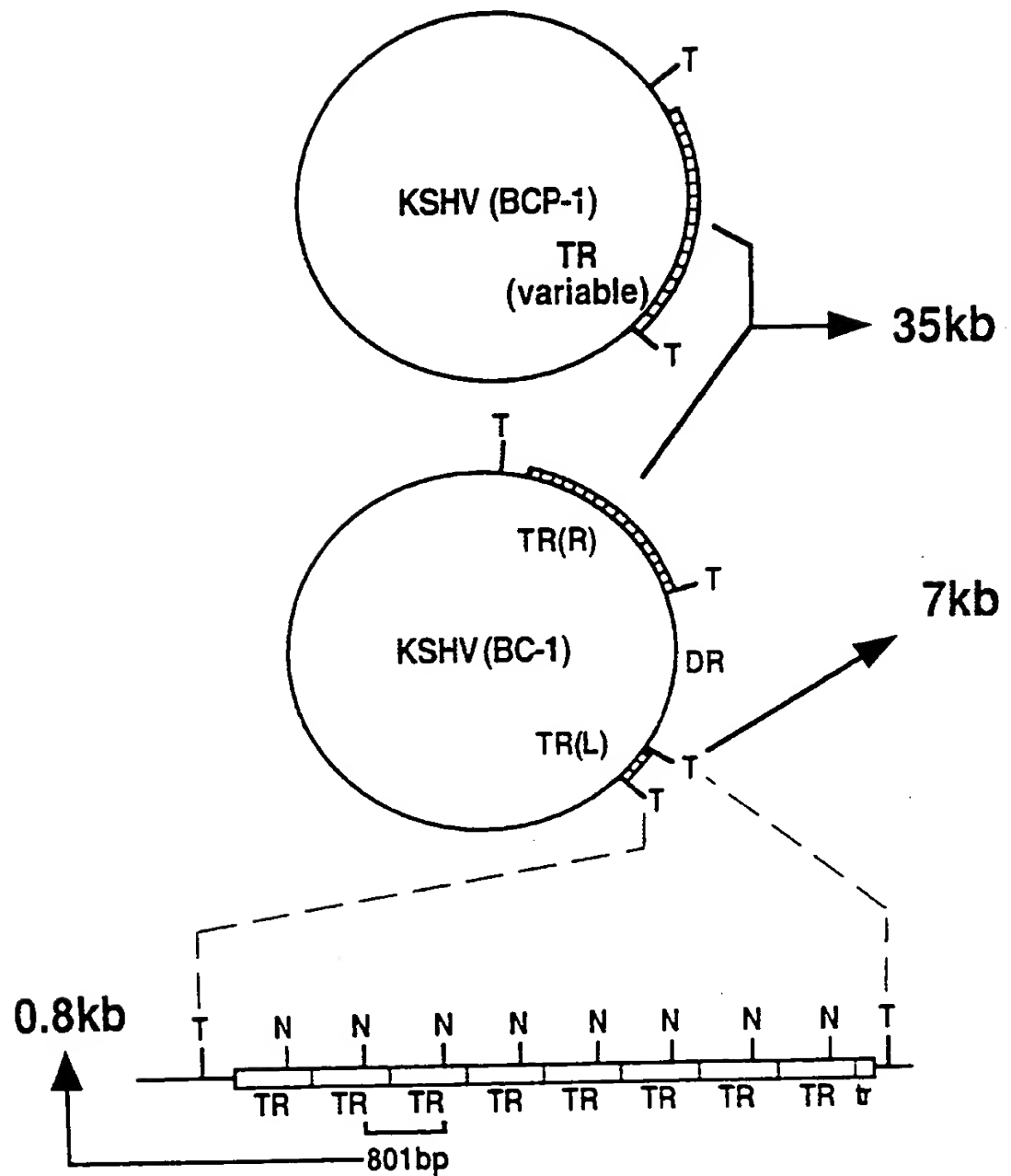


0.8kb



4/15

FIG. 2C



5/15

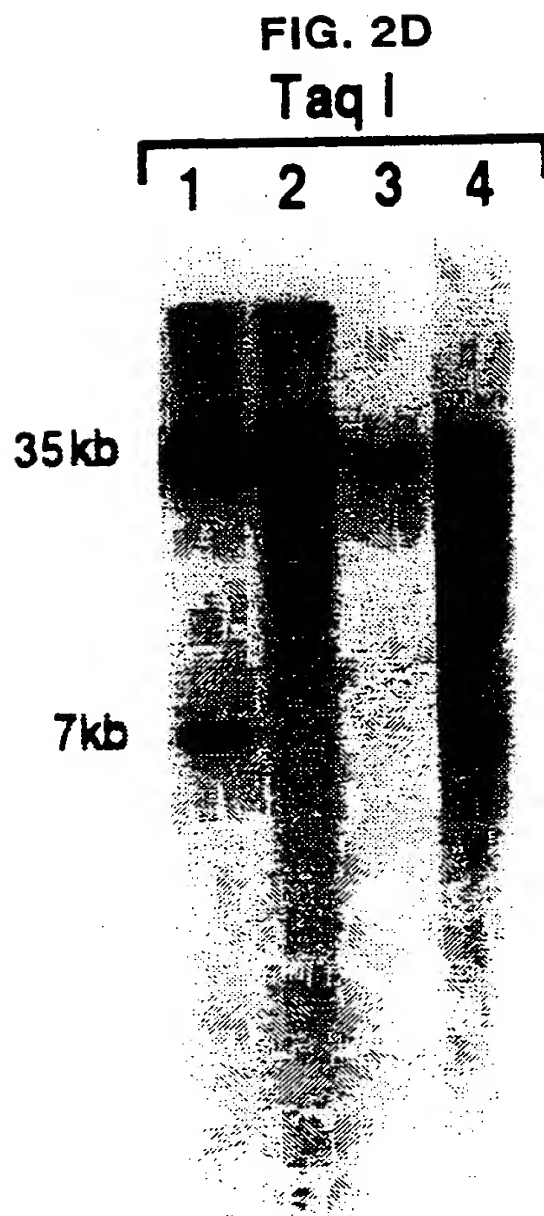




FIG. 3B

|       |       |        |    |     |      |            |    |     |    |    |    |    |    |    |    |    |     |
|-------|-------|--------|----|-----|------|------------|----|-----|----|----|----|----|----|----|----|----|-----|
| VIL6  | MCWF  | KLWSLL | LV | --- | GSLL | VSGTRGKLPD | AP | -   | EF | ED | KL | -  | LI | QR | LN | WM | 50  |
| huIL6 | MNSF  | ST9APG | PV | AP  | SL   | GL         | LL | VLP | AA | FP | AP | V  | PP | GE | DS | KD | VS  |
| VIL6  | LWV   | ---    | I  | DEC | FRD  | ---        | -  | CY  | RT | GI | CK | IL | EP | AA | IF | HL | 100 |
| huIL6 | ERID  | KQIRYI | LD | GI  | SA   | LR         | KE | TC  | NK | SN | MC | ES | SK | EA | LA | EN | 100 |
| VIL6  | CGLI  | GFNE   | TS | CL  | KL   | AD         | GE | FE  | EV | LF | KK | FL | TE | FG | KS | VI | 150 |
| huIL6 | CPQS  | GFNE   | ET | CL  | VK   | II         | TC | LL  | EE | EV | Y  | LE | Y  | LN | Q  | AA | 150 |
| VIL6  | GWDI  | QEEL   | NK | LT  | KT   | HY         | SP | PK  | FD | RG | LL | GR | LQ | GL | KY | W  | 200 |
| huIL6 | IQFL  | QKKA   | KN | LD  | AI   | TT         | PD | PT  | TN | AS | LL | TK | LQ | AA | QN | QW | 200 |
| VIL6  | FAGQ  | AVR    | VL | D   | S    | IP         | DV | TP  | DV | H  | DK |    |    |    |    |    |     |
| huIL6 | FLOSS | LR     | AL | -   | -    | -          | -  | -   | -  | -  | -  | -  | -  | -  | -  | -  | -   |

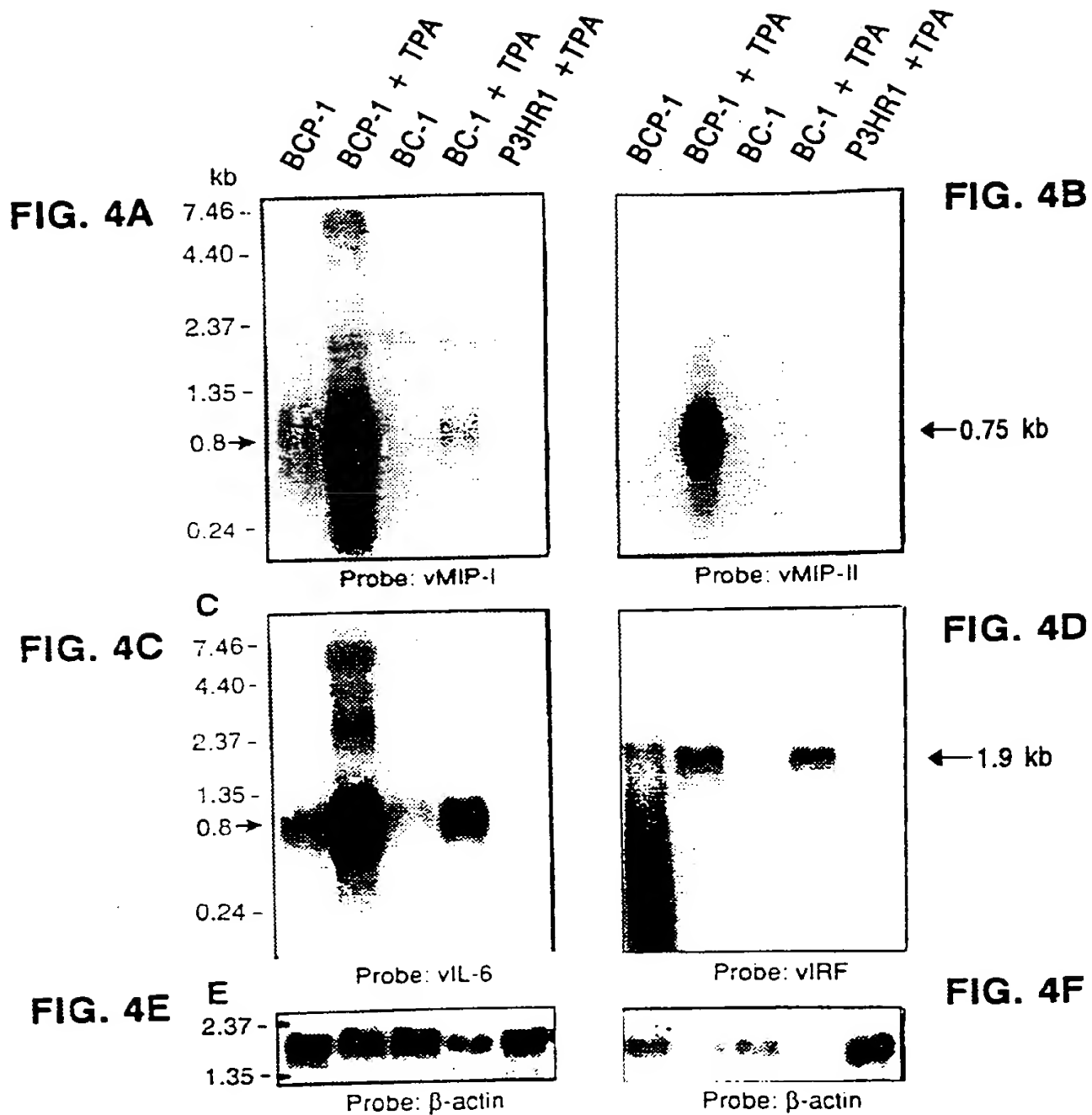
FIG. 3C-1

|          |            |            |            |            |            |     |
|----------|------------|------------|------------|------------|------------|-----|
| VIRP     | MDPGQRPNPF | GAPGAIPKKP | CLSQGSPGTS | GSGAPCDEPS | RSESPGEGPS | 50  |
| huISGF3γ |            |            |            |            |            |     |
| huICSBP  |            |            |            |            |            |     |
| VIRP     | GTGGSAAAGD | ITRQAVVAAI | TEWSRTRQLR | ISTGASEGKA | SISKD      | 100 |
| huISGF3γ |            |            |            | MASGRARCTR | KLRN       |     |
| huICSBP  |            |            |            | MCDRNGGR   | -LRQWLI    |     |
| VIRP     | NSGKFPQV   | EW         | EDERTFR    | EWRRD      | GELGV      | 150 |
| huISGF3γ | ESGQFPQV   | CKW        | DDTAKT     | REDQDA     | AFFK       |     |
| huICSBP  | DSSMY      | PGLIK      | ENEEKSM    | NQEVDA     | ASIFK      |     |
| VIRP     | VDASPKG    | TRG        | RRMLA      | GISQD      | GHHFL      | 200 |
| huISGF3γ | EGD        | TGGPA      | VW         | GRMD       | VAE        |     |
| huICSBP  | EGDKA      | EPATW      | KTRLR      | SOLDI      | SEPYK      |     |
| VIRP     | QC         | VECGV      | VAG        | RL         | QEGFPS     | 250 |
| huISGF3γ | SGQ        | PGTQ       | KVP        | NCTLS      | PSV        |     |
| huICSBP  | KCK        | LGVA       | TAG        | EP         | SVDD       |     |
| VIRP     | CTTA       | EGQE       | AV         | QEV        | TD         | 300 |
| huISGF3γ | SGGA       | VHSD       | IG         | FQGD       | QRS        |     |
| huICSBP  | PPDA       | CRSQ       | LL         | YDA        | HHSA       |     |

FIG. 3C-2

|          |              |              |             |             |             |       |
|----------|--------------|--------------|-------------|-------------|-------------|-------|
| VIRF     | FIRMYYNGEQ   | VHELLETTTSQS | GCRISALRR   | PAVHYCAVG   | SPGQVWLP    | N 350 |
| huISGF3Y | LLTFIYNGRV   | VGEAQVQB LD  | -CRLVAE--   | -P--SGS-ES  | SMEQVLFPP   | KP    |
| huICSBP  | VISFYYGK L   | VGOATTTCPE   | GCRLSLSQPG  | LPGTKLYGPE  | GLELVRFP    | -P    |
| VIRF     | VPNLACEIAK   | RELCDTLDAC   | AKGILLTSSC  | NGIFCVCYHN  | GPVHFIGNT   | V 400 |
| huISGF3Y | GPLEPT--QR   | --LLSQL      | ERGILVASNP  | RGLEFVQRLCP | IPISWNAPQA  |       |
| huICSBP  | ADTIPSERQR   | QVTRKLFCHL   | ERGVLHSSR-  | QGVFVKRRLCQ | GRVFCVVVV   | V     |
| VIRF     | PPDSGPP LLLP | QCKPTRIENP   | NTFLVGLAN-  | -S--PLPAP   | SHVTCPLVKL  | 450   |
| huISGF3Y | PPGPGPHLLP   | SNECVELEFRT  | AYFCRDLLVRY | FQGLGPPPKF  | QVTLNFWEEES |       |
| huICSBP  | VCKORPNKLE   | RDEVVQVFD T  | SQFFRELQQF  | YNSQGRLLPDG | RVVLCFCGEEF |       |
| VIRF     | WLGKPVAVGK   | LEPHAPSP--   | RDFAAARCSNF | SDACVLEIM   | PKPLWDAMQ   | - 500 |
| huISGF3Y | HGSSHTPQNL   | ITVKMEQAFA   | RYLLEQTPEQ  | QAAILSLV--  | --- --      |       |
| huICSBP  | PDMAPLRSKL   | ILVQIEQLYV   | RQLAEEAGKS  | CGAGSVMQAP  | EEPPPDQVFR  |       |
| VIRF     | -----        | -----        | -----       | -----       | -----       |       |
| huISGF3Y | -----        | -----        | -----       | -----       | -----       |       |
| huICSBP  | MPPDICASHQ   | RSFFRENQOI   | TV          | -----       | -----       |       |

10/15



11/15

FIG. 5A

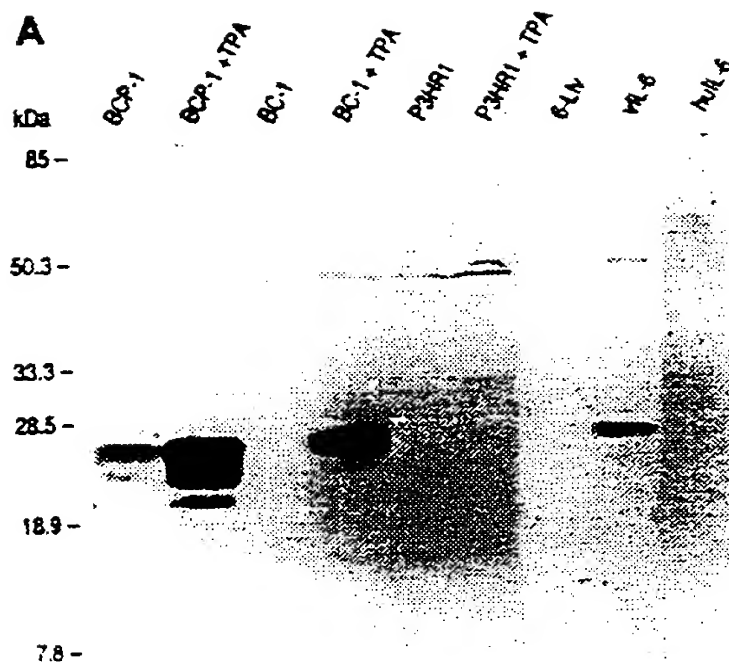
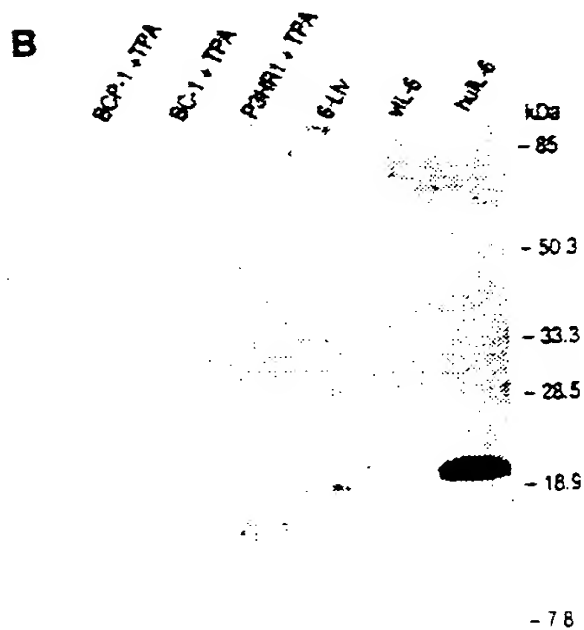


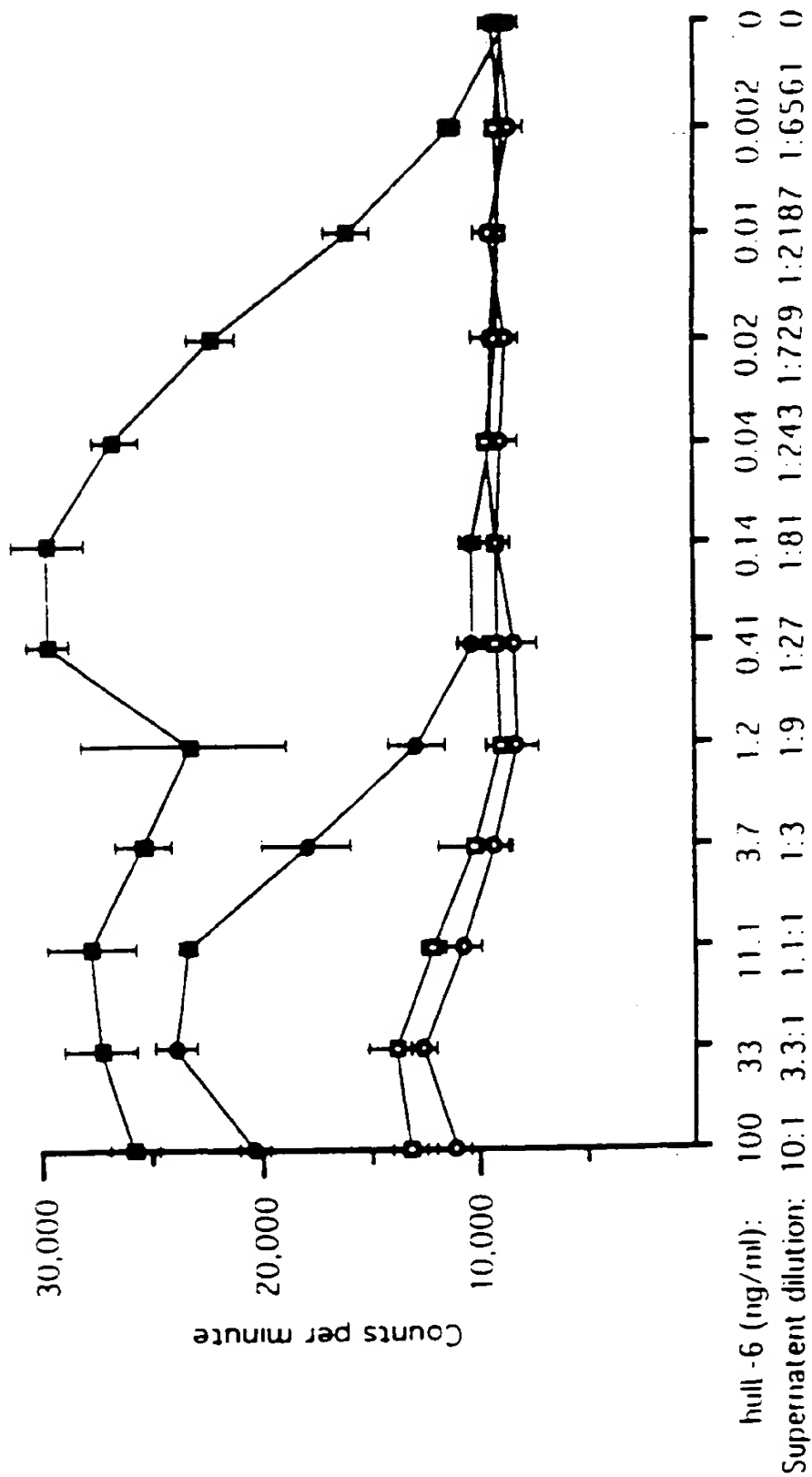
FIG. 5B





12/15

FIG. 6



13/15

FIG. 7A



FIG. 7B



FIG. 7C



FIG. 7D



FIG. 7E

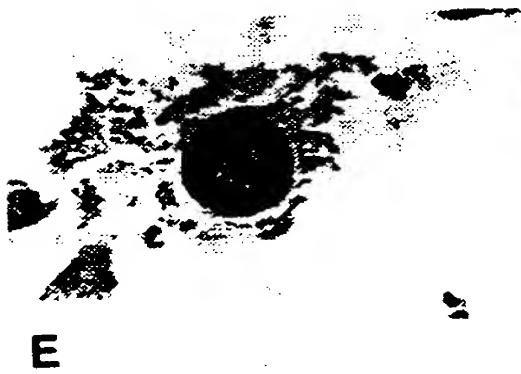
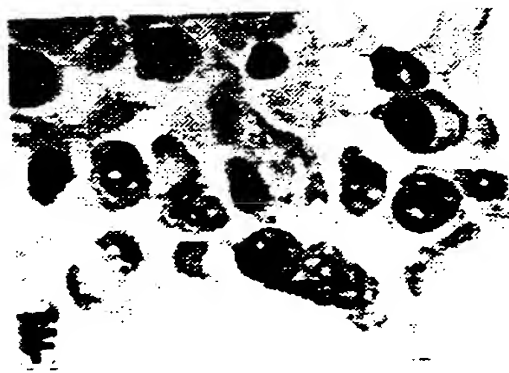


FIG. 7F



14/15

FIG. 8A

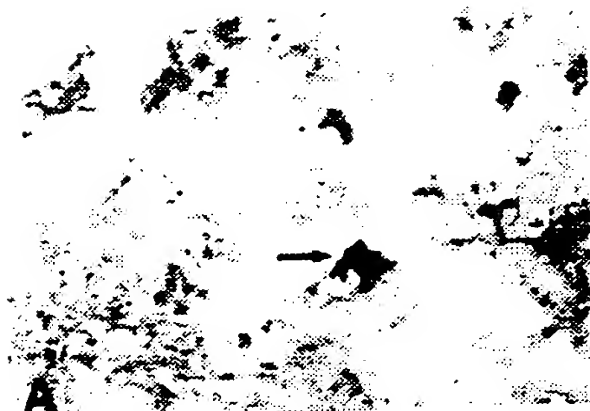


FIG. 8B



C



FIG. 8C

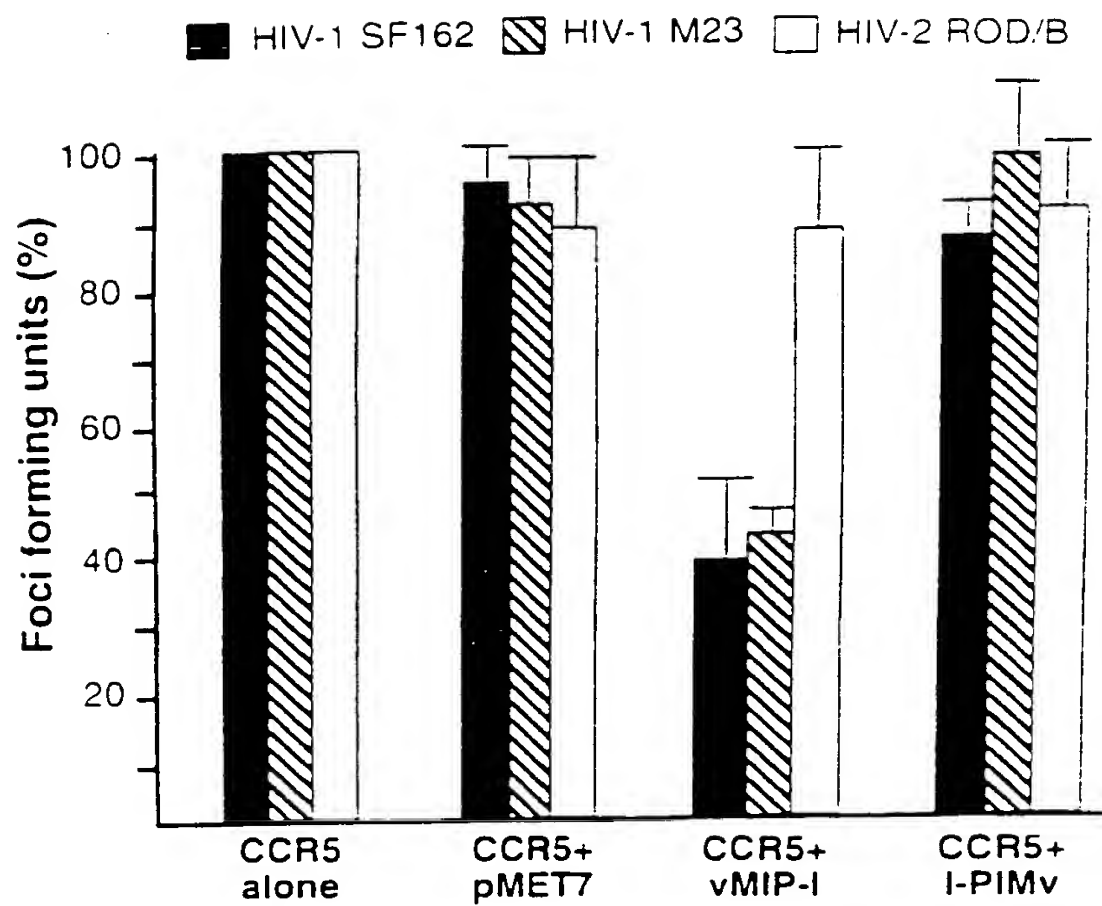
D



FIG. 8D

15/15

FIG. 9



# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/01442

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : Please See Extra Sheet.  
US CL : Please See Extra Sheet.  
According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols):

U.S. : 536/23.72, 24.3; 424/199.1, 204.1, 230.1, 184.1, 130.1, 229.1; 435/5, 6, 7.1, 41, 240.23, 235.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS, WPIDS, AIDSLINE. Search terms: HHV-8, herpesvirus 8, human herpes 8, Kaposi's sarcoma, ks?

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
|-----------|--|-----------------------|
| Y         | Chang et al. Identification of Herpes-Like virus DNA Sequences in AIDS-Associated Kaposi's Sarcoma. Science, 16 December 1994, Vol. 266, pages 1865-1869, see entire document.   | 1-12, 19, 25-28       |
| Y         | Memar et al. Human herpesvirus-8: detection of novel herpesvirus-like DNA sequences in Kaposi's sarcoma and other lesions. Journal of Molecular Medicine, 10 October 1995, Vol. 73, pages 603-609, see entire document.                  | 1-12, 19, 25-28       |
| Y         | Moore et al. Detection of Herpesvirus-like DNA Sequences in Kaposi's Sarcoma in patients with and those without HIV infection. The New England Journal of Medicine. 04 May 1995, Vol. 332, No. 18, pages 1181-1185, see entire document. | 1-12, 19, 25, and 26  |

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

|   |     |  |
|---|-----|--|
| * Special categories of cited documents:  | *T  | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |
| *A* documents defining the general state of the art which is not considered to be of particular relevance   | *X* | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |
| *E* earlier documents published on or after the international filing date   | *Y* | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | *G* | document member of the same patent family  |
| *O* document referring to an oral disclosure, use, exhibition or other means  |     |  |
| *P* document published prior to the international filing date but later than the priority date claimed  |     |  |

Date of the actual completion of the international search  
28 APRIL 1997

Date of mailing of the international search report  
12 MAY 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231  
Facsimile No. (703) 305-3230

Authorized officer  
ALI R. SALIM

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/01442

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No. |
|-----------|---|-----------------------|
| Y         | Cesarman et al. In Vitro Establishment and Characterization of Two Acquired Immunodeficiency Syndrome-Related Lymphoma Cell Lines (BC-1 and BC-2) Containing Kaposi's Sarcoma-Associated Herpesvirus-Like (KSHV) DNA Sequences. Blood, 01 October 1995, Vol. 86, No. 7, pages 2708-2714, see entire document. | 1-12, 19, 25-28       |

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/01442

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:  
1-12, 19, 25, 26, and 27-28
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☒  
☐

- The additional search fees were accompanied by the applicant's protest.  
No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/01442

**A. CLASSIFICATION OF SUBJECT MATTER:**  
IPC (6):

C07H 21/04; A61K 39/12, 39/245; C12Q 1/70, 1/68; G01N 33/53; C12P 1/00; C12N 5/00; A61K 39/00, 39/395

**A. CLASSIFICATION OF SUBJECT MATTER:**  
US CL :

536/23.72, 24.3; 424/199.1, 204.1, 230.1, 184.1, 130.1, 229.1; 435/5, 6, 7.1, 41, 240.23, 235.1